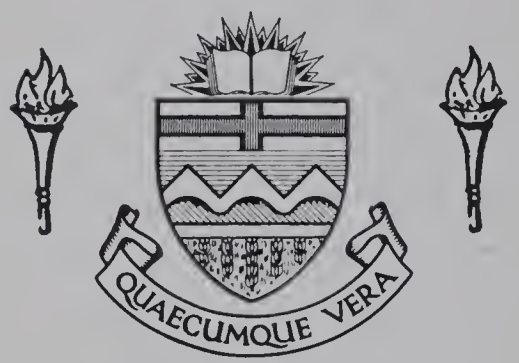


THE UNIVERSITY OF ALBERTA

RELEASE FORM

NAME OF AUTHOR Han-Yong You

TITLE OF THESIS An Acoustic and Perceptual Study of English Fricatives

DEGREE FOR WHICH THESIS WAS PRESENTED Master of Science

YEAR THIS DEGREE GRANTED 1979

THE UNIVERSITY OF ALBERTA

AN ACOUSTIC AND PERCEPTUAL STUDY OF ENGLISH FRICATIVES

by

HAN-YONG YOU

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE

OF MASTER OF SCIENCE

IN

SPEECH PRODUCTION AND PERCEPTION

DEPARTMENT OF LINGUISTICS

EDMONTON, ALBERTA

SPRING, 1979

THE UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research, for acceptance, a thesis entitled ___An acoustic and perceptual study of English fricatives_____

submitted by _____Han-Yong You_____

in partial fulfilment of the requirements for the degree of

Master of Science

in Speech Production and Perception.

ABSTRACT

The discrimination of the signal properties of labial and dental fricatives has long been a controversial issue in the study of English consonants. In an attempt to solve this long-standing problem, as well as examining general issues concerning fricatives, the present study analyzed three sources of acoustic information: the duration, the overall intensity, and the spectral configuration, of the frication portion of fricative consonants.

By means of discriminant function analyses, it was shown that the more or less independent effects due to VOICE and due to PLACE can be extracted from the spectral composite. Thus, a generalized voicing effect on spectral configuration was obtained.

The analyses concering PLACE indicated that English fricatives constitute a hierachical structure of PLACE categories, in which [alveolar] and [alveo-palatal] are grouped into a "back" superordinate PLACE class, and [labial] and [dental] into a "front" superordinate PLACE class. These two superordinate classes were well distinguished by any source of information considered in this study. The distinction between back fricatives appeared to be possible simply by detecting the spectral pole of the

signals.

From a perceptual experiment, it was concluded that the frication portion of front fricatives also has sufficient information with which a PLACE identification can be fairly well achieved. The results of the further analyses of front fricatives suggested the conclusion that a pole-zero detecting strategy is a plausible perceptual process for PLACE identification cf front fricatives.

ACKNOWLEDGEMENTS

I am especially indebted to all of the members of my committee for their kind and assiduous diagnoses of my logic, ideas and speculations painted with non-native accents dumped onto the first draft of this thesis shortly before the Christmas holidays of 1978 winter. I would like to thank my supervisor, Dr. Anton Rozsypal who suggested the original topic for this study, and has helped me understand the fundamentals cf acoustic theories. I have to express a deep gratitude to Dr. Terrance M. Nearey for his valuable guidance in building the general theoretical framework for this study as well as for his kind advice in conducting experiments and analyses.

I thank Dr. John T. Hogan who always made me feel welcome to knock at his door to discuss about any kind of topics. I am grateful to Dr. William J. Baker for his concise and clear explanations concerning any statistical problems I have ever had. Special thanks are forwarded to Dr. Kyril T. Holden, my external examiner, for the sincere interest in my research he has shown since the excerpt was read on the fourth ACOL conference at Banff, and for the warm encouragement and keen criticisms he gave me. I would also like to acknowledge Dr. Kellog Willson's aid in my accessing to the subject pool of Psychology Department for

this study. I want to remark particularly that this thesis would not have been completed smoothly but for the silent help of Mr. Dale S. Stevenson, my fellow student in psycho-acoustic group. This study owed tremendously to his Alligator system program and MTS-PDP communication program.

I am very happy to acknowledge that I am greatly indebted to my wife, Bong-Hee Choi, who has also been my colleage in the field of linguistics since Seoul National University. Compared with the spiritual encouragement and heartful trust she has always provided me, her immense editorial helps for preparing this thesis were rather trivial.

I thank the Department of Linguisitcs which supported my graduate study for three years through graduate assistantships.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

NOTATIONS AND SYMBOLS

In the present thesis, the names of all phonetic or phonological features, the names of all the acoustic or physiological parameters for speech production, and all the statistical factors appear in upper case letters. Specific values of phonetic or phonological features are always in square brackets. Whenever it was required to denote tokens rather than an abstract signal or feature, I.P.A. symbols delimited by slashes are used. Abbreviations or mnemonics are used only after an explicit specification unless they are the ones widely accepted by tradition. All statistical symbols are underscored.

CHAPTER ONE

INTRODUCTION

This chapter sets the theoretical framework and general background in which the present study is to be set, and specifies the object of the study together with some fundamental assumptions taken by the present study as working hypotheses.

1. FEATURE FRAMEWORK FOR SPEECH DESCRIPTION

To facilitate a descriptive study of language, it is generally assumed that speech is a sequence of discrete entities i.e., certain phonological units in terms of which a sentence is encoded. Halle (1964:325) states concerning this:

> Almost every insight gained by modern linguistics from Grimm's Law to Jakobson's distinctive features depends crucially on the assumption that speech is a sequence of discrete entities. In view of this fact many linguists have been willing to postulate the existence of discrete entities in speech event while accepting as true the assertion of instrumental phoneticians that there are no procedures for isolating these entities.

Even if the existence of certain phonological units is taken for granted, the question of how to define these units remains. Modern phonological and phonetic research has normally assumed that the phoneme-size segment is to be further analyzed into feature-size elements. The present study will thus assume with Lieberman (1968:3) that "the phonological feature is the minimal linguistic unit in terms of which speech is coded". The present study will further assume that an utterance is basically the realization of a stream of phonological features rather than a string of segments. As to the nature of the phonetic features, the position of Lieberman will be adopted:

> A phonological feature must have both acoustic and articulatory correlates because speech must be produced by the human vocal tract, and must be transmitted through the air. Discussions of whether the acoustic or articulatory level is more "basic" often are really discussions of whether it is easier to obtain and classify acoustic or articulatory data. (1968:3)

Since there is no direct empirical access to speech codes, that is, to phonological features, a feature system similar to the traditional PLACE by MANNER by VOICE description will be assumed in the present study.

This section is devoted to a brief sketch of these features and their traditional definition. Within the traditional descriptive framework, speech was essentially thought of as an output of the interaction of phonation and

articulation (cf. Minifie, Hixon & William, 1973:127-209). This view of speech production which was originated by Sanskrit grammarians (cf. Allen, 1953) has been developed and quantitatively formulated by modern phoneticians to establish the source-filter theory of speech production, which will be reviewed in the following section to provide the present study with a theoretical base.

2. SOURCE-FILTER THEORY OF CONSONANT PRODUCTION

According to Lieberman (1977:35f), the source filter theory of speech production was first proposed by Johannes Müller (1840). Recently, it has been put on a quantitative basis by Chiba & Kajiyama (1941) and Fant (1960). Modern models of consonant production have been successfully implemented using quantal parameters corresponding closely to some traditional features.

Phonation --- VOICE

Jones (1960:19) indicated that vocal folds are capable of acting in much the same way as the lips during speech. According to him, vocal folds are in a breathing position when they are wide apart as in breathing, and that they are in voicing position when they are drawn close together so

that they can vibrate when air is forced through them. Furthermore, he also defined that the "voiced" sounds are the sounds produced with vocal folds in voicing position, and the "breathed" or "voiceless" sounds are those produced with vocal folds in breathing position (for more precise definition of glottal mode, cf. van den Berg, 1968). There is a peculiarity to be noticed in the glottal ccnstriction, which characterizes VOICE factor. Since vocal folds are quite flexible, rapid "close-open" alternations of vocal folds due to Bernouilli effect take place when a fast air stream flows through the constricted vocal folds. Thus voicing results in the emission of a quasi-periodic train of air puffs into the supraglottal passage. This emmission of air puffs serves as the source of a voiced sound.

VOICE is determined by many articulatory parameters involved in controlling the state of glottal opening (or the mode of glottal vibration) and the configuration of larynx. There are a few distinct possible modes of VOICE such as [voiceless], [whisper], [voiced], [creak], [breathy voicing or murmur] and [creaky voice] (refer to Catford, 1968:318-321; van den Berg, 1968:291-301; Ladefoged, 1971:6-9; Huh, 1968:20; Lieberman, 1968:14-18). However, since VOICE is realized only as either [voiced] or [voiceless] in normal speech acts of English, we will assume that VOICE is a dichotomous factor.

Articulation --- MANNER and PLACE

According to modern acoustic models of speech production (especially of consonant production), two parameters related to the configuration of vocal tract[1] are involved in articulation. The two parameters which are essentially the most important in determining the acoustic outputs are: the size S and the location L of the minimal cross-sectional area (i.e., the maximal constriction area).[2] It is obvious that S, which is highly correlated with the height of an articulator, is a generalized parameter which determines the extent of narrowing of the air-passage at some critical point while L is the one which determines the location where the narrowing of air-passage takes place.

MANNER

If S is greater than a certain value[3] the output is a vocalic[4] where the airflow is more or less a smooth laminar stream provided the subglottal air pressure is kept normal. If S is zero, there can be no airflow through the vocal

---

[1] By "vocal tract", like Stevens & House (1955) or Chiba & Kajiyama (1941), we will narrowly mean the airway from glottis through oral cavity up to lips only, excluding nasal cavity

[2] Stevens & House intended to develop the parameters to explain acoustic ccrrelates of vowels. However, we find that all oral sounds can be explained in terms of these paramenters.

[3] Stevens (1971:1188) reports that the value is $S \geq 0.3$ cm^2.

[4] Refer to the definitions of "vocalic" by Jakobson, Fant & Halle (1969: 18-19) and by Chomsky & Halle (1968: 176-177)

tract until the blockage is released to produce a stop or an affricate according to the release speed dS/dt. If S is within a certain range[1] such that a turbulent airflow occurs, then a fricative is produced. Thus, S can be regarded as a parameter whose value varies on a continuous scale between the two extremes: the maximum of 2 cm^2 (Stevens, 1971:1188) to produce low vowel, and the minimum of 0 cm^2 to produce silence for a stop consonant.

As an articulatory parameter, the term "APERTURE" introduced by de Saussure (1969:70-76) as early as the beginning of this century, seems to be the most applicable to S. APERTURE, though conceptually narrower, is closely related to the traditional term "MANNER". Note for example that APERTURE is defined independently of NASALITY which is frequently dealt with as a distinct MANNER class in traditional phonetics. Ladefoged (1967:11) indicated that the magnitude of the subglottal air pressure (henceforth SUBGLOPRES) which determines the volume velocity of the airstream (Stevens, 1971:1181; Lieberman, 1968:22-37) interacts with APERTURE in characterizing the supraglottal source. In addition, in the traditional description of the MANNER class "fricative", auditory concepts such as "hissing sound" (Jones, 1960:47) were frequently used. But, APERTURE is defined in absolute articulatory terms, and undoubtedly

---

[1] Stevens (1971:1187) reports that the range of S is $0.05\ cm^2 \leq S \leq 0.2\ cm^2$.

it is a major determinant of the feature MANNER. However, besides APERTURE, some other factors such as SUBGLOPRES or GLOTTAL-CONFIGURATION are also involved in determining the presence of frication noise. Considering the narrowness of the term, APERTURE, the traditional MANNER distinction appears to be better suited to the characterization of fricative sounds than APERTURE alone.

PLACE

The location L of minimal cross-sectional area in the vocal tract can also take an arbitrary point on the longitudinal axis of the continuous vocal tract delimited by the lips and glottis at each end. This parameter L mainly determines two acoustically relevant factors:

(1) -(i) the configurations of the front cavity and the back cavity due to the division by L (cf. Heinz & Stevens, 1961), and

(ii) the shape and length of the constriction (for "shape", see Flanagan, 1972:189).

Flanagan (1972:72-77) described in detail the functions which determine poles and zeroes of spectral shapes in terms of the lengths of the front cavity and the back cavity. We can produce different speech sounds with various spectral peaks and valleys even though these are excited by, roughly speaking, the same supraglottal source by means of shifting

L back and forth along the longitudinal axis of the vocal tract while keeping S constant as a certain determined value. In contrast, we can produce different speech sounds with similar spectral peaks and valleys but excited by various supraglottal sources by means of adjusting S along its continuum while fixing L at a certain determined position. Therefore, the articulatory event L is a parameter determining mostly the locations of poles and zeroes of speech sounds in their spectra. We will henceforth reserve the term LOCATION for this parameter.

But, it has been noted that pole and zero configuration of a spectrum is not determined by LOCATION alone. Nearey (1977:25) pointed out that the raising or lowering of the larynx (henceforth "LARYNGEALITY") and the protruding or retracting of the lips (henceforth "LABIALITY") can play a role in determining the transfer characteristics of the vocal tract. Minifie et al. (1973:269) also noted that the configuration of lip opening is also relevant to pole and zero configuration of a spectrum. Minifie et al. (1973:259) stated, concerning the shape of the cross-sectional area, that a circular orifice is a more effective noise generator than an eliptical orifice of the same area. Evidently, front cavity can be lengthened by adding lip protrusion and/or moving the location of the minimal cross-sectional area posteriorly in the vocal tract. For example, the acoustic difference between the fricatives, /s/ and /ʃ/, may

be due to the differences in the shape of lip opening as well. The fact that /ʃ/ has a lower primary formant frequency than /s/, appears not less attributable to the effect of the lip-protrusion for an articulation of /ʃ/ or /ʒ/ than to the effect of a palatalization of constriction location.[1] Thus, it seems more descriptive to call /s/ unrounded alveolar fricative, and /ʃ/ rounded alveo-palatal fricative[2] as far as English fricatives are concerned. Thus, LABIALITY[3] and LARYNGEALITY are the factors which are articulatorily orthogonal to LOCATION, but which interact with one another to produce a similar acoustic effect. Therefore, PLACE may in fact be an articulatory complex which includes such minor factors as LABIALITY and LARYNGEALITY as well as LOCATION. LOCATION and LABIALITY are not fully crossed in English, and both factors are determinants of the configuration of the vocal tract. Hence, we will regard PLACE as a simple articulatory factor with four categorical levels: [labial], [dental], [alveolar] and [alveo-palatal]. Through a series of experiments testing perceptual confusions of English consonants, Miller

---

[1] Ladefoged (1975:279) states concerning this that "Labialization <is> a secondary articulation in which lip rounding is added to a sound, as in English /ʃ/".

[2] It is interesting to note that a naive Korean speaker would think that the difference between the strings /sa/ and /ʃa/ does not lie in PLACE of consonants, but in LABIALITY of vowels since Korean orthography treats [rounded] of /ʃ/ as a property of the following vowel sequence rather than one of the preceeding consonant.

[3] Addressing the measurement of vertical and horizontal distance of lip-opening, Ladefoged (1975:257-267) discerns LABIALITY and ROUNDNESS.

& Nicely (1955) and Lindblom and Serpa-Leitao (1969) discovered that PLACE identification was much more prone to confusion than MANNER or VOICE identification. Their experiments clearly demonstrated that most perceptual confusions in consonant identification took place between PLACEs within a correct identification of MANNER and VOICE. This result indicates that the acoustic cues for PLACE identification are very likely to be more subtle than those for VOICE or MANNER idenfication.

Acoustic aspects of PLACE difference

It has already been stated that PLACE differences are related primarily to different configurations of front and back cavities of the vocal tract. Since the configurations of these cavities play an important role in determining poles and zeroes of resonance characteristics, a PLACE difference is strongly expected to be realized at the acoustic level primarily in terms of different formant/antiformant frequencies. Minifie et al. (1973:260) stated that "the closer the sound source is to the lip opening, the higher will be the natural resonant frequency of the vocal tract". This indicates that the front cavity functions as major resonance cavity for fricative formants. As to zeroes, Flanagan (1972:73) stated that "zeroes occur at frequencies where the impedence looking back from the source (toward the glottis) is infinite". The same idea was

expressed by Fant (1962:10) as follows:

> zeroes will appear at approximately the same frequencies as the poles representing the resonances of the back cavities.

In a very simple model, front cavity may be thought of as a tube open at one end, and back cavity as a tube closed at both ends. Therefore, pole frequencies are predicted by

(2) $Fn = c(2N - 1)/4Lf$; $N = 1, 2, 3, \ldots$

where Fn is the frequency of Nth pole, Lf is the length of front cavity, c the velocity of sound. Zero frequencies are predicted by

(3) $Fm = cM/(2Lb)$; $M = 1, 2, 3, \ldots$

where Fm is the frequency of Mth zero and Lb is the length of back cavity.

Independence of phonation and articulation

Not only are the articulatory parameters containing voicing and supraglottal constriction essentially independent of each other, but according to the source-filter theory of speech production, their acoustic consequenses are also essentially orthogonal.[1] Thus, if we

---

[1] There is a slight qualification of this with respect to fricatives since the presence of the supraglottal noise source is affected by glottal modification of ariflow. This point is pursued in the next chapter. It has only minor practical consequences.

combine both features, namely MANNER and VOICE assuming that both are dichotomous, then we get a full combination of the two constrictions as shown in Table 1. It is shown that a voiced sound is produced when there is a glottal constriction and a voiceless sound is produced when there is

Table 1. Combinatorial table of Constrictions.

		Supraglottal Constriction	
		yes	no
Glottal Constr.	voicing position	Voiced Obstruent	Vocalic
	breathing position	Voiceless Obstruent	Aspiration (i.e., /h/)

no glottal constriction, and an obstruent (ignoring its NASALITY) is produced when there is a certain supraglottal noise-generating constriction and a vocalic is produced elsewhere.

CHAPTER TWO

BACKGROUND OF THE PROBLEM

This chapter is devoted to the review of previous studies of fricative consonants. It has three sections: the first section concerns general properties of the specific speech signal [fricative], i.e., a fixed value of MANNER, with some consideration of the mechanism of its production, while the other two sections concern distinctions of various values of VOICE and PLACE with MANNER held constant as [fricative].

1. GENERAL PROPERTIES OF FRICATIVES

Aerodynamic aspects of fricative production

Like most speech sounds, fricatives are also best defined by describing their production mechanism. Jones (1960:47) states that

> Fricative consonants are formed by a narrowing of the air-passage at some point to such an extent that, the escaping air, expelled by pressure from lungs, produces

audible friction, i.e., some kind of hissing sound.

This definition attempts to capture the basic articulatory dimensions of fricatives, that is, the fact that fricatives are produced when APERTURE is within a certain range no matter where the constriction is located. Heinz & Stevens (1961:589) described this as follows:

> Air is forced through this constriction at high velocity, and turbulent flow occurs in the vicinity of the constriction and possibly also at the teeth[1]. Noise is generated as a result of the turbulent flow, and this noise acts as excitation for the acoustic tube that forms the constriction and for the cavities anterior to the constriction. There may also be some acoustical coupling through the constriction to the rear cavities.

Stevens (1971) states that the aerodynamic factor which determines source characteristics is the magnitude of pressure drop Pd caused by energy losses due to any constriction in the air-passage. Since Pd is given by

(4) $Pd = kU^2/S^2$,

where U is the volume velocity and k is an appropriate constant (Stevens, 1971:1181), the major factor determining Pd is the magunitude of S provided U is assumed constant during speech.[2] A dimensionless parameter Reynold's

---

[1] Flanagan (1972:53) noted that a turbulent airflow can also be generated by directing an air jet across an obstacle or sharp edge.

[2] But note that U can be affected by both changes in glottal configuration and/or subglottal pressure variation.

number Re given by

(5) $Re = vw/k$,

where v is the particle velocity, w the effective width of the air-passage, and k an appropriate constant determined by velocity and density of the gas (Minifie et al., 1973:258) is an important reference in understanding turbulent streaming and noise generation in fricatives. Flanagan (1972:55) stated that the magnitude of the sound pressure of a noise Pr is approximately determined by

(6) $Pr = k(Re^2 - Rc^2)$,

where Rc is a critical value for a certain shape of constriction, and k is a constant. Considering that $v = U/S$ and that S can be proportional to w^2, by substituting (5), (6) can be rewritten as

(7) $Pr = jU^2/S - kRc^2$,

where k and j are appropriate constants. Formula (7) indicates that Pr increases (above some threshold kRc^2) when S becomes smaller provided the volume velocity U is held constant, and that Pr increases when U becomes greater provided the constriction area S is held constant.

Stevens (1971:1183) classified turbulent noise into two groups: "frication noise" due to a supraglottal turbulence, and "aspiration noise" due to a glottal turbulence. But he

added that the difference of the two groups involves not only the region where a frication is generated but also how Pr increases beyond the critical value so that a frication can occur. A frication noise (i.e., a supraglottal noise) is undoubtedly due to a decreased S. But an aspiration noise (i.e., glottal fricative) is rather due to an increased volume velocity U. An aspiration noise /h/ is produced when vocal folds abduct closer to their breathing position so that a vibration for voicing cannot take place. Stevens (1971:1188) thus argues that

> The average airflow for voiced vowel sounds produced by adult male speakers is normally in the range 100 - 200 cm^3/sec. ... <but> there are occasions during speech production when the vocal tract is in a vowel-like configuration but the glottis is spread and there may be no vocal-cord vibration. The airflow under these conditions is ccnsiderably greater than it is during normally voiced vowel sounds, and is generally in the range 500 - 1500 cm^3/sec. These airflow conditions occur for voiceless vowels, for /h/ sound, and for the aspiration that follows the release of certain stop consonants in English.

In the present study, our interest will be limited to frication noises only. Thus, we will henceforth mean by "fricatives" only those sounds produced by an excitation of the vocal tract cavities by a turbulent noise source due to any <u>supraglottal constriction</u> (whether or not there is an active glottal constriction), thus excluding /h/ sound.

Acoustic characteristics of fricatives

The differences of signal type due to different MANNER are readily evident in common spectrographic displays. The signal [fricative], produced by an APERTURE value less than [vocalic] but greater than [stop] or [affricate], exhibits an intermediate characteristics of signal pattern between these two extremes. The signal property of "aperiodicity" of [fricative] is shared by [stop] or [affricate], while the signal property of "continuity" of [fricative] is shared by [vocalic].

As early as 1947, Potter, Kopp & Kopp (1966:35) observed in their spectrum analyses that [fricative] is easily identified by the "irregular fill". Fant (1962:13) noted that [fricative] "is recognized by a high frequency noise area in the spectrum". Denes & Pinson (1963:134) stated that

> Unquestionably, we distinguished fricatives from all other sounds by the 'hissy' noise the turbulent air stream produces which shows up on spectrograms as a fuzzy segment.

As previous acoustic studies concerning fricatives consistently indicate, [fricative] is identified by the existence of continuant spectral energy spread widely over the high frequency bands (beyond 2 kHz) (for detailed description of general spectral configuration of fricatives, see Minifie et al., 1973:258-272; Ladefoged, 1975:179).

There have been a number of previous studies which attempted to explore the distinctiveness of PLACE or VOICE of fricatives. These will be reviewed in the following sections. Each of the studies was conducted along one of the following three tracks:

(8) -(i) static analysis of long-term frequency spectra of some frication sections,

(ii) dynamic analysis of short-term frequency spectra of some time-dependent fricative signal portion, or

(iii) absolute or relative measurement of some prosodic properties, e.g., duration, overall intensity, etc.

2. VOICE DISTINCTION OF FRICATIVES

This section is devoted to the review of a number of studies which attempted to find acoustic correlates of VOICE in English fricative consonants, by means of acoustic measurements or psychoacoustic experiments.

Hughes and Halle's spectral analysis

Hughes & Halle's (1956) study is one of the earliest attempts to present a quantitative description of the

acoustic properties of English fricatives. A few isolated words containing fricatives were recorded by two male and one female speakers. By means of a trigger-tone-inserting technique, a 50 msec long section of frication was gated from the recorded signals (the criteria for locating this section is not clear) and subjected to spectral analysis, which gave frequency density spectra within the range of 0.3 - 10 kHz. Hughes & Halle noticed that in the region above 1 kHz, the spectra of voiced fricatives and voiceless fricatives do not differ appreciably, and that voiced fricatives often have "a very strong component in the region below 700 Hz" (they called this "voicing component") while "this region is never prominent" in voiceless fricatives. But they argued that

> since it was not true that a voiced fricative has always a voicing component, the distinction at acoustic level between voiced and voiceless fricatives is not necessarily made on the basis of a low-frequency component in the spectrum.

Jassem's spectral analysis

Jassem's (1965) spectral analyses of fricatives were conducted in a much more rigorous way. He sampled CV or CVC type nonsense syllables containing [fricative] from one native speaker each of English, Swedish and Polish in the context of the extreme vowels /a/, /i/, and /u/. From the sampled data, each 80 msec long section ending 10 - 20 msec

before the end of frication portion was taken for a spectral analysis in the frequency range of 0 - 9 kHz. In his analysis, Jassem reported that VOICE was easily identified in all three languages by the spectral information alone. Jassem claimed that

> In each pair of <homorganic> voiced-voiceless fricatives there is a cross-over point in the spectral envelope. Below this the voiced sound has more energy, and above it the pressure level is higher in the voiceless sound. The energy below the cross-over point (which occurs somewhere between 600 and 2300 Hz) in voiced fricatives is almost entirely due to the voice source.

Besides this, he noticed that total energy level was also found to be different according to VOICE. He did not attempt to explain why, but nevertheless he reported that a voiced fricative has less overall noise-source energy than a homorganic voiceless fricative.

Rabiner's claim for two components of voicing effects

In an experiment of speech synthesis by a terminal-analog synthesizer, Rabiner (1967a:20-21; 1967b:824-825) claimed that voicing is manifested in fricatives as the superposition of an amplitude-modulated supraglottal noise and the glottal voice source. Rabiner also argued that "unvoiced component of voiced fricatives" was the component due to the supraglottal noise, and "voiced component of voiced fricatives", the component due to the glottal source. However, Rabiner (1967b:824-825) pointed out that the

unvoiced component of voiced fricative, (i.e., the lower formants of voiced fricatives) was remarkably less pronounced than predicted by his model.

Cole & Cooper's test of duration effect

Cole & Cooper (1975) conducted a perceptual experiment to investigate the effects of durations of the fricative portion and of the contextual vocalic portion, on the perception of VOICE identification of some English fricatives and affricates.

They obtained samples of /fa/, /sa/ and /tʃa/ syllables from a male speaker. By means of a tape-slicing technique, each whole frication portion was divided into six equi-distant subsections. They produced six new signals by recording the consecutive playbacks of the frication portion and the vocalic portion with the last subsection of the frication taken off one by one. These six new stimuli, differing in their frication duration, were presented to subjects who were forced to decide on VOICE of the fricatives or affricates. The results showed that truncation of four or five subsections of the frication led subjects to make more responses for [voiced] than for [voiceless]. On this account, Cole & Cooper carried out a test of the analysis of variance (ANOVA) to examine the significance level of physical duration difference between

[voiced] and [voiceless] tokens. The ANOVA results indicated that a voiced fricative clearly had a shorter frication portion than its voiceless cognate.

Massaro & Cohen's test of voice-onset time effect

Massaro & Cohen (1976) hypothesized that voice-onset time (VOT) would play an important role in VOICE identification in fricatives as well as in stops. In order to test this hypothesis, Massaro & Cohen synthesized /si/ and /zi/ by means of a digital-synthesis technique controlling the duration of VOT[1] so that stimuli with four different VOTs (70, 90, 110, and 130 msec long) were generated. These stimuli were presented to five subjects who were forced to decide if the stimuli sounded more like /si/ or /zi/. The results indicated that their responses moved more toward /si/ from /zi/ when VOT was increased.

Discussion and summary of VOICE identification

First of all, it was found that the existence of "voicing component" in low frequency area in the spectra of voiced fricatives was not always detected. This fact

---

[1] In most studies of stops, VOT was measured as the interval between the burst of noise and the onset of voicing. As to fricatives, however, it is quite difficult to find a stable reference time point. But as Massaro & Cohen indicated, since the duration of the synthesized frication was held constant as 200 msec, this issue was avoided in their experiment.

suggests that the frication section of fricatives might not be so uniform throughout its duration as is generally assumed. Especially in the case of a long voiced fricative in which a voicing starts relatively late, the results of a spectral analysis could differ substantially depending on which portion of the frication section is to be analyzed. It is worthwhile to note that the only fricative in which Hughes & Halle could not find a voicing component was /z/, which turned out to be much longer than other fricatives in the present study (see Chapter Four). But it should be pointed out that Hughes & Halle's gating-window was only 50 msec long while Jassem's was 80 msec long. Hence, it is suspected that Hughes & Halle might have set their window prior to the onset of voicing. Such a time-dependent factor in speech analyses deserves special attention. It underscores the necessity of considering the dynamic aspects of speech sounds and that speech signals must be regarded as strictly time-dependent. The appearence of such studies as Massaro & Cohen which was concerned with voice onset time coincides with this new trend in the study of speech sounds which regards an utterance as a stream of features, rather than as a string of segments, where each segment is a block of a few concurrent features.

The question of how the spectral data are to be represented remains open. Hughes & Halle (1956) adopted a within-signal calibration method which normalizes the

spectra so that, for instance, the maximum level is always valued as 0 dB. In contrast, Jassem (1965) adopted a cross-signal calibration method which takes some arbitrary apparatus-dependent value as a reference for the spectral levels. If there exists a significant difference in intrinsic intensity (see Lehiste and Peterson, 1959:429) between signals, or if volume control during the recording process was not consistent, the absolute-calibration method could be seriously misleading. On the other hand, the within-signal calibration method involves assumptions which are not necessarily perceptually correct since overall intensity is lost. There is no a priori knowledge which of the plausible alternatives is the most suitable for the description of our speech behavior.

It should also be realized that an experimenter must be very careful with his experimental design especially in a perceptual test. For instance, Cole & Cooper's perceptual experiment using signal truncation technique resulted in a conclusion which was compatible with their ANOVA test of frication duration mentioned above. However, it must be pointed out that they happened to manipulate a potentially relevant factor in their experiment which they did not consider in fact. The voicing of a vowel after a voiceless fricative does not begin until well after the frication ends resulting in a considerable length of "intensity minimum" period (Peterson & Lehiste, 1960:696; and LaRiviere, Winitz

& Herriman, 1975:615). But in a voiced fricative, voicing starts much earlier so that there is no abrupt change in the amplitude envelope from frication portion to vocalic portion. In their experiment, the splicing technique of Cole & Cooper ultimately resulted in an effectual smoothing of the amplitude envelope between the twc portions by cutting off the intensity minimum part from the frication portion of voiceless fricatives. There is little doubt that this effect must have made the remaining portions sound more voiced-like.

Looking over all the previous literature dealing with VOICE identificatiоn, we find that the reported acoustic correlates (and presumably acoustic cues as well) for [voiced] were mostly:

(9) -(i) the existence of the so-called voicing component (below about 700 Hz),

(ii) less overall energy in the non-voicing component regions (above about 700 Hz),

(iii) shorter duration of frication portion,

(iv) earlier onset of voicing, and

(v) amplitude modulation of supraglottal noise by voice fundamental.

3. PLACE DISTINCTION OF FRICATIVES

This section reviews a number of important previous studies which attempted to discover experimentally acoustic correlates of, or perceptual cues for PLACE feature of English fricatives.

Hughes and Halle's spectral analysis

The study of Hughes & Halle (1956) was one of the earliest works on PLACE identification of fricatives. It should be noted that dental fricatives (i.e., /θ/ and /ð/) were not considered in their analyses. They derived some classification criteria for distinguishing the three PLACEs on the basis of the comparisons of the spectral levels of various frequency regions (see Hughes & Halle, 1956:308) and concluded:

(10)-(i) Peak frequencies of each PLACE of fricatives within each subject[1] showed a consistent order, that is, [alveo-palatal] < [alveolar] < [labial],

(ii) [alveo-palatal] rarely has a strong concentration of energy above 4 kHz, while [alveolar] and [labial] do so,

(iii) [alveolar] has more concentrated energy in lower

---

[1] LaRiviere (1974) investigated speaker differentiation with fricatives.

half band (2.15 - 6.00 kHz), while [labial] has more evenly distributed energy, and

(iv) [alveo-palatal] has the narrowest energy peak.

Harris' test of effects of contextual vowel

Harris (1958) attempted to investigate the effect of the succeeding contextual vowels on PLACE identification of English fricatives. She edited real speech sounds by means of a tape-slicing technique to get 64 artificially coupled fricative-vowel strings. The same process was done saparately with voiced and voiceless groups. She combined the frication portion obtained from 4 different PLACEs (with VOICE held constant) to each of all 4 different contexts of the same vowel. This routine was repeated with each of the 4 vowels, namely /i/, /e/, /o/ and /u/. These stimuli were presented to 22 subjects who judged PLACE of the stimuli.

From this perceptual test, she concluded that frication portions of labials and dentals appeared perceptually indistinguishable, and that these two groups were distinguished by the cues from their contextual vowels, while others were well identified independently of contextual vowels. But interestingly enough, it was also found that labials and dentals could be also well distinguished by their frication portions if the contextual vowels were rounded.

Strevens' spectral analysis

Strevens' (1960) study considered the overall intensity of the frication portion and the width of the frequency range over which the random energy is spread as candidates for PLACE identification cues. Tokens of voiceless fricatives including some non-English sounds were recorded by 13 well-trained phoneticians. To achieve better spectral shapes, subjects were asked to produce fricatives of extraordinarily long duration (1000 msec).[1] In a visual examination of spectrograms covering the range of 0 - 8 kHz, he reported that:

(11)-(i) more overall intensity level was detected in more posterior fricatives,

(ii) while the frequency range of energy concentration of front fricatives was relatively wide (5 - 6 kHz wide), that of back fricatives was relatively narrow (3 - 4 kHz wide), and

(iii) while the energy peak of front fricatives was below 2 kHz,and that of backs was above 2 kHz.[2]

---

[1] For reasons that will be clear in the following discussions, it is convenient to refer to [labial] and [dental] together as "front fricatives" and to [alveolar] and [alveo-palatal] as "back fricatives".

[2] Notice that item (iii) is contrary to the reports of some others (e.g., Heinz & Stevens, 1961).

Heinz and Stevens' circuit simulation of fricatives

Heinz & Stevens (1961) limited the scope of their study to PLACE identification of voiceless fricatives. At first, they considered the aerodynamic mechanism of production of voiceless fricatives, and provided a two-pole-one-zero model which is conceptually analogous to a simplified vocal tract configuration for voiceless fricatives. The spectra were simulated by an electric circuit, and compared with those obtained from real speech sounds (the data provided by Hughes & Halle, 1956). Finally, they conducted a perceptual experiment using synthesized stimuli in order to test the validity of their predictions. In the synthesis process, however, they adopted an even more simplified model[1] with one pole and no zero to get a reasonable approximation of the outputs of two-pole-one-zero model. The simplified model was controlled by the frequency and bandwidth of the single pole. The analysis of subjects' responses indicated that more anterior PLACE responses resulted as pole frequency increased. This experiment also indicated that [dental] and [labial] could not be distinguished from each other with the information obtained from spectral configuration alone. They carried out another experiment in which the relative intensity level of frication portion (with reference to that of the vowel portion), and the locus

---

[1] The effectiveness of this simplified model was confirmed by Fujisaki & Kunisaki's (1978) study of Japanese fricatives.

of second formant transition of the contextual vowel were concurrently considered.

Their study concluded that

(12)-(i) pole frequency is relevant to PLACE identification,
(ii) overall intensity level of the frication portion with reference to that of the contextual vowel is also relevant to PLACE identification, and
(iii) second formant transition is also relevant to PLACE identification, particularly with [dental] and [labial].

Delattre, Liberman and Cooper's test of transition effect

Delattre, Liberman & Cooper (1964) investigated some acoustic cues for fricative PLACE identification, especially between [dental] and [labial] in terms of some contextual parameters. Delattre et al. chose voiced fricatives only, assuming that cues for PLACE should not be affected by VOICE factor. There were two main reasons why voiced fricatives were selected:

(13)-(i) lower overall intensity of a voiced fricative may reduce the information from spectral cues, and
(ii) longer transition period of voiced fricatives would provide better experimental control.

By means of a pattern-playback technique, they produced synthesized stimuli from hand-painted spectrograms. Each stimulus had a contextual vowel with their formant transitions completely controlled. The result supported the conclusion that [dental] and [labial] can be distinguished by different loci of the second and third formant transitions.

Jassem's spectral analysis

Jassem (1965) measured the levels and loci of four formants, and compared the formants with one ancther. In order to obtain some parameters by which maximum discrimination of the PLACEs of fricatives could be achieved, he defined for English two acoustic features:

(14)-(i) SPREADNESS: if F4 - F3 ≥ 1.8 kHz
then [+spread]
elsewhere [-spread], and

(ii) HIGH-FREQUENCY-EMPHASIS (HFQEMP):
if ([+spread] and L4 - L2 > 0)
or ([-spread] and L3 - L2 > 0)
then [+hfqemp]
elsewhere [-hfqemp],

where F_i indicates the frequency of the ith formant and L_i, the dB level of the ith formant. In terms of these two

acoustic features, English fricatives were discriminated as indicated in Table 2. His study was noteworthy in that he

Table 2. Feature specification of fricatives by Jassem (Notice that this table fails to distinguish [dental] and [labial]).

	labial	dental	alveolar	alveo-pal
SPREAD	+	+	+	-
HFQEMP	-	-	+	-

not only proposed some explicitly defined acoustic features to distinguish among fricatives, but also relied on relative measures rather than absolute values of loci or levels of formants and antiformants. However, in spite of his complex and disjunctively defined criteria, [labial] and [dental] were not distinguished from each other.

LaRiviere, Winitz and Herriman's test of transition effect

LaRiviere, Winitz and Herriman (1975) carried out an experiment to investigate the effect of the information from succeeding vocalic portion in a design similar to that of Harris' (1958). LaRiviere et al. divided fricative-vowel syllables into three portions: frication, transition, and vowel. They experimentally controlled the length of the

transition into five different levels (from 0 to 250 msec) to produce various stimuli in one experiment. In another experiment, they controlled the type of stimuli by eliminating or preserving the transition or the vocalic portion after the frication. The results of their perceptual experiments indicated that the back and the front fricatives are distinguished in terms of the degree of their contextual dependency. They reported that vowel information does not play a significant role in perception of back fricatives, while it does so for front fricatives, especially when followed by the contextual vowel /i/. It was also reported that the inter-speaker variation of spectral peaks was greater in the case of front fricatives than in back fricatives.

Fujisaki & Kunisaki's model for Japanese fricatives

Fujisaki & Kunisaki (1978) recorded 60 CV and VCV type words containing /s/ or /ʃ/, and lowpass-filtered them at 9.6 kHz for digital sampling. The frication segments were extracted by a 50 msec Hanning window (refer to Rabiner & Gold, 1975:75-102) and were converted into logarithmic power spectra over the frequency range of 0 - 10 kHz. The range from 0.3 - 5.0 kHz was taken, and was divided into 24 bands of equal width each of which was represented by the average level within the band. Using the least mean-square error criterion, they determined the frequencies of poles and

zeroes of various models (one-pole model, two-pole model, one-pole-one-zero model, and two-pole-one-zero model), and measured the rms error value of each model. They reported that the models with a zero are much superior to those without a zero for the approximation of /ʃ/, but the difference is diminished for the approximation of /s/. On the basis of this analysis, an optimum linear discriminant function for each model was derived to separate /s/ and /ʃ/, and percentage of correct discrimination by means of those functions were examined. The results indicated that 100 percent recognition was possible with models including a zero while errors occurred in the case of models without zeroes.

Fujisaki & Kunisaki then tested perceptual effectiveness of the models. They synthesized speech by computer simulation of a terminal-analog synthesizer. The outputs were presented to subjects who categorized the stimuli. The results indicated that the two-pole-one-zero model and the one-pole-no-zero model produced the best approximations. Through a naturalness judgement test, they demonstrated that the two-pole-one-zero model and the one-pole-no-zero models were superior to other models, but they were not significantly different from each other. Though they could not explain the reason for the discrepancy between the results obtained by an analysis and by perceptual tests, and even if it is not clear whether or not

the models proposed by them will apply to English fricatives in which there are two more categories, Fujisaki & Kunisaki's work was highly instructive in that it suggested a way of evaluating models for fricative simulation.

Discussion and summary of PLACE identification

It was noticed that the findings discovered by a number of investigators concerning PLACE identification of fricatives are in general quite consistent. But a remarkable controversy was observed concerning front fricatives. Strevens (1960) claimed that front fricatives have a peak below 2 kHz, while Heinz & Stevens (1961) reported that a peak is found in a much higher frequency area, namely beyond 8 kHz. However, it should be remembered that Stevens also pointed out that the bandwidths of front fricatives were considerably wider than those of other groups, while the overall intensity of front fricatives is the lowest. On this account, we can reasonably infer that the spectrum of frontal fricatives are more or less flat and depressed and the peak is nct so prominent. The acoustic correlates of fricative PLACEs reported by previous research are thus summarized as follows:

(15)-(i) front fricatives are distinguished from back fricatives by a wide and flat spectrum,

(ii) [dental] and [labial] (i.e. front fricatives) are poorly distinguished from each other by their

static spectral configuration,

(iii) [dental] and [labial] can be distinguished from each other primarily by the formant transition into the following vowel, and

(iv) [alveolar] is identified by a smooth peak in high frequency area and [alveo-palatal] is identified by a sharp peak in middle frequency area.

CHAPTER THREE

QUANTITATIVE ANALYSIS OF FRICATIVES

This chapter describes the data used in the present study and reports a series of quantititative analyses designed to discover acoustic correlates of PLACE and VOICE differences in English fricatives by means of various statistical tests. The present study analyzes three sources of information: duration of frication, overall intensity of frication, and spectral configuration of frication. The first section describes some relevant matters concerning the choice of parameters for the data and their measurements. The second section describes the analyses of the duration and the overall intensity of the frication portion.[1] The last section is devoted to the analyses of spectral properties of the frication portion which is expected to carry most of the linguistically significant information.

---

[1] For convenience, we will denote the information of the overall intensity and the duration of the frication portion by "prosodic information", though this constitutes an extension of the ordinary usage of this term.

1. DESCRIPTION AND MEASUREMENT OF DATA

Speakers

Three twenty-year-old male speakers were recorded. The speakers were all native Edmontonians.

Materials

In preparing the experimental material, the four fully crossed factors shcwn below were controlled at the following levels (Refer to Appendix A for details):

(16)-(i) four PLACEs of articulation: [labial], [dental], [alveolar] and [alveo-palatal],

(ii) two VOICEs of glottal source: [voiced] and [voiceless],

(iii) four contextual vowels (VOWELs) in post-consonantal context: /i/, /æ/, /ʌ/ and /u/,

(iv) three SPEAKERs: speaker 1, speaker 2, and speaker 3, and

(v) two speaker repetitions, i.e., two repetitions of each syllable produced by each speaker (SPKREPs)[1].

---

[1] Unless otherwise specified, SPKREP is treated as experimental replication in the present study.

In order to enhance the naturalness of the reading of the nonsense material, the controlled monosyllable was embedded into a carrier sentence:[1]

(17) Please, say that /F-W-d/ again.

where fricative F and vowel W were controlled as described. The speakers were asked to put a stress on the controlled material to make the acoustic effect of the controlled factors more distinctively realized. A voiced dental stop /d/ was selected as the final consonant of the controlled syllable for the follcwing reasons:

(18)-(i) a stop is easily segmentized from W, and

(ii) final dental stops have relatively minor coarticulation effects on the spectral values of the initial consonant and following vowel formant transition (see Broad & Fertig, 1970).

Apparatus

The instruments below were used in this study, and their essential technical specifications were, according to the manufacturer, as follows:

(19)-(i) Microphone - Sennheiser MD 421N

- Frequency response: 30 to 17000 Hz ± 5 dB, with

---

[1] This idea was adopted from that of Lehiste & Peterson (1958:431).

5 dB rise between 3 kHz and 10 kHz

- Sensitivity: 0.2 mV/micro-bar (for 1000Hz)
- Directionality characteristics: cardioid.

(ii) Tape recorder - TEAC A-7030

- Frequency response: 50 to 15000 Hz ± 2 dB
- Speed: 7.5 ips
- Signal-to-noise ratio: 58 dB

(iii) Audio-frequency filter - Frøkjaer-Jensen type 400 (for anti-aliasing)

- Slope of frequency response: 36 dB/oct

(iv) Minicomputer - PDP-12A

- Word length: 12 bits
- Analog/Digital converter: 10 bits
- Digital/Analog converter: 10 bits
- Operating system: (a) OS/8; (b) Alligator[1]

Data acquisition and preprocessing

Recording

Recording was conducted in a sound-insulated recording room. Only the left channel was employed so as to avoid any possible crosstalk. The tempo of speech was regulated by

---

[1] Alligator is an operating system program for psychoacoustic experimentation executable on a PDP-12 or similar computers, written in OS/8 PAL-12D assembly language. For detailed description of the system, refer to Stevenson & Stephens (1978a; 1978b).

repeatedly presenting to the speakers a control sentence pre-recorded by a trained speaker. While a speaker was uttering his sentence, the experimenter kept watching the VU meter of the tape-recorder ensuring that the input level to the recorder was always around or below 0 dB so as to avoid a signal distortion.

Digital gating and segmentation

This procedure was executed by an interactive program (henceforth "gating prcgram") written in Alligator. Two major steps were cyclically executed with each recorded utterance: the digital gating[1] of an audio-signal, and segmentation of the frication and the vocalic portions. This procedure is schematically represented in Figure 1. The sampling step converted analog speech signals recorded on the audio-tape into an array of digitized numbers in the work area of Alligator. Prior to digitizing by an A/D converter, the audio-signal from the tape recorder was first bandpass-filtered to eliminate 60 Hz hum and possible speech frequency components above 8 kHz (one half of the sampling rate of 16 kHz). Thus, the anti-aliasing filter was set to a bandpass range of 68 to 6800 Hz, and attenuation slope of

---

[1] For detailed theoretical and technical discussions concerning digital gating of speech signals, refer to Rozsypal (1976).

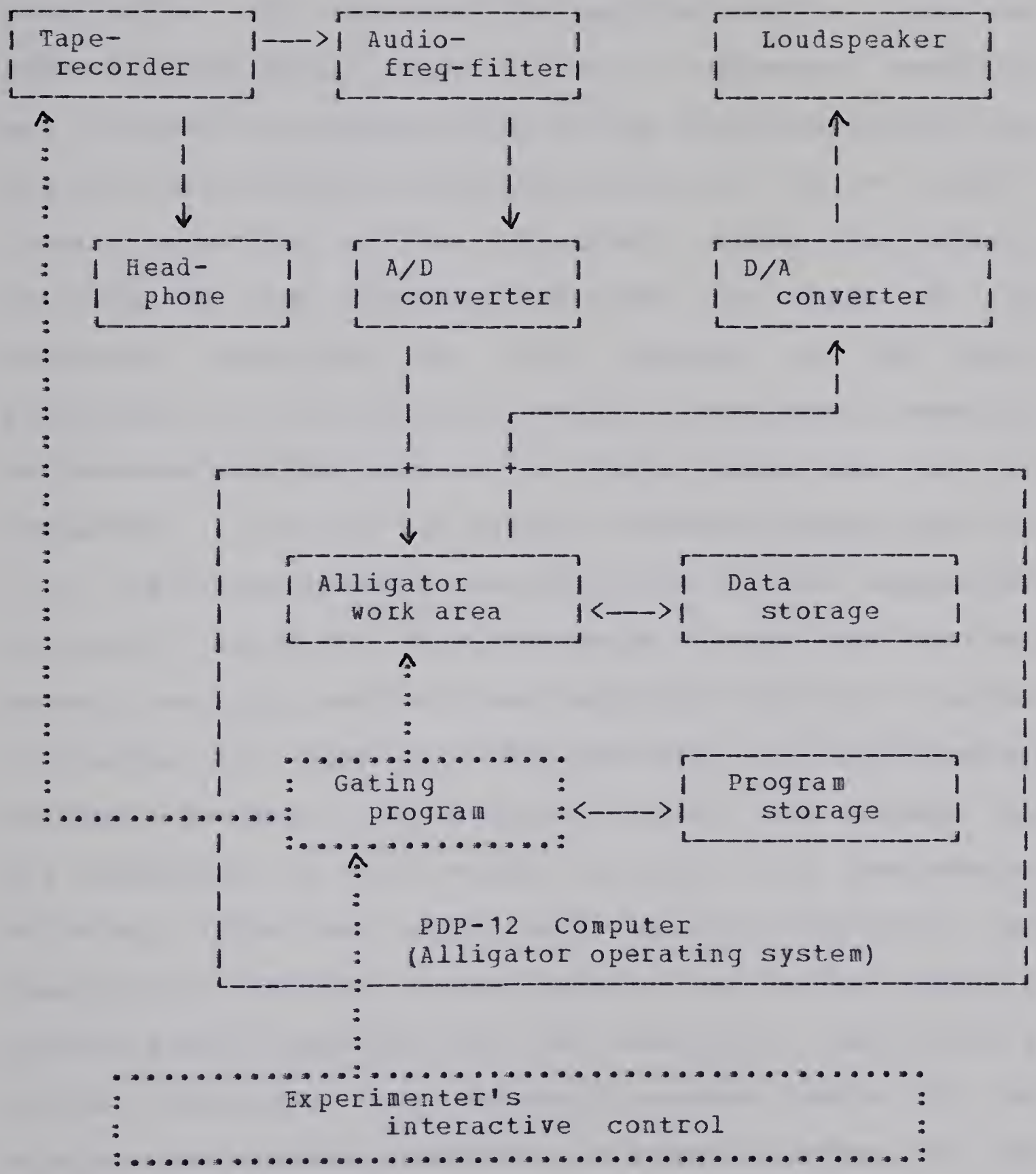


Figure 1. Block diagram of digital gating and segmentation.
*Solid arrows indicate signal flows; dotted arrows, control flows; solid boxes, devices; and doted boxes, ccntrollers.

36 dB/oct.[1] While the recorder was playing back the utterance, the signal was auditorily monitored by the experimenter who initiated the sampling routine. Care was taken to avoid signal clipping. Once a successful sampling was obtained, the segmentation of the frication portion and the vocalic portion was carried out with the aid of audio-visual inspection of the digitized signal. The primary criteria for the segmentations were the shape of the amplitude envelopes and the presence of the voice periodicity of the contextual vowel. A pronounced intensity minimum was observed between the burst noise due to the explosion of /t/ in the carrier sentence "Please say that ..." and the noise due to the frication of the controlled fricative consonant. Thus, the point at which the distance between the upper and the lower amplitude envelopes reached a minimum was taken as the beginning of the frication portion. The end of the frication portion was assumed as the beginning of the vocalic portion. In the case of voiceless fricatives which were usually demarcated by unambiguous intensity minimum points, locating the boundary between the two portions was not difficult. But, for a voiced fricative, the intensity minimum point in the waveform was extremely difficult to locate because of the voice modulation. Thus, for voiced fricatives, the appearance of higher frequency periodicity due to the second

---

[1] A pilot study indicated that 8 PLACE-by-VOICE categories of fricatives from one speaker could be well identified by a discriminant function analysis under these conditions.

formant of the following vowel was regarded as the beginning of the vocalic portion.

Measurement of data

The audio-signals digitized and stored by the gating program were processed by a FORTRAN program (henceforth "measurement program") run on the PDP-12 computer so that appropriate quantitative data could be obtained. The measurement program measured the durations of the frication and the vocalic portions, and performed a spectral analysis on the frication portion. First, the DC bias was removed. Then, the duration and the overall intensity of each segment were measured. The 64 msec segment of the frication portion with maximum intensity was selected for spectral analyses. This segment was then multiplied by a Hamming window (Rabiner & Gold, 1975:75-102). When the frication portion was shorter than 64 msec, the signal was padded with trailing zeroes. The discrete amplitude spectra were calculated using a routine based on the fast Fourier transformation (FFT) and converted to dB values. Each cell of the array represented a 15.625 Hz bandwidth. Sets of sixteen consecutive 15.625 Hz bands were averaged so that the entire spectrum was divided into 32 non-overlapping

frequency bands (henceforth BINs) each 250 Hz wide.[1] A conversion table between BIN number and frequency is provided in Appendix A. The data obtained by the measurement program on the PDP-12 were transferred to the Amdahl 470V/6 university computer for statistical analyses.

2. ANALYSIS OF OVERALL INTENSITY AND DURATION

This section concerns the contribution of the duration and the overall intensity of the frication portion to the identification of PLACE and VOICE factors. For each measurement, an ANOVA is first conducted to examine the significance level of the effects due to the controlled factors and their interactions. A Newman-Keuls procedure for testing significance of differences between individual means for significant factors is carried out. Finally, a series of discriminant function analyses (DFA's) are conducted to explore the discriminability of the data in terms of these factors.

---

[1] Considering the average fundamental frequency of a male voice, the width of 250 Hz may be too wide to capture the information of every harmonic of the voice fundamental. This process may suppress some information defined as "voice component of voiced fricatives" by Rabiner (1967a).

Test of significance of duration difference

The aim of this test was to discover if the duration of the frication portion (or any parameter directly dependent on it) is significantly related to PLACE and/or VOICE differences. Two alternative measures were considered: the absolute frication duration (henceforth "DUR") and the ratio of DUR to the duration of the contextual vowel W (henceforth "RDUR"). For each of the two alternative variables, a 4-way ANOVA was conducted in which the factors SPEAKER, VOWEL, PLACE, and VOICE were fully crossed with two repetitions. The results of the ANOVA's as shown in Table 3, indicate that even though RDUR was less speaker-dependent than DUR, it was much more context-dependent. Moreover, most of the variance of the duration of W was due to the vowel /æ/ which was longer than others by the ratio of 3:2. Nevertheless, the value of DUR associated with this vowel was found to be the shortest. Thus, RDUR appeared not to contribute to a context-normalization. Thus, it was decided to take DUR as the relevant measure of fricative duration in this analysis.

The main effect of PLACE and the interaction effect of SPEAKER-by-PLACE were significant at .001 and .01 levels, respectively. This indicates that the duration of the frication portion is significantly associated with PLACE although the magnitude of the effect can vary across

Table 3. F-ratios of ANOVA's for duration of the frication portion (DUR) and for ratio of frication duration to vowel duration (RDUR).

Source	df		F for	F for
	num.	denom.	DUR	RDUR
S	2	96	3.57*	2.00
W	3	6	2.24	13.92**
P	3	6	23.77**	41.63***
V	1	2	6.84	10.59
SW	6	96	4.98***	2.67*
SP	6	96	3.30**	1.48
WP	9	18	3.34*	2.78*
SV	2	96	4.04*	4.67*
WV	3	6	3.96	3.60
PV	3	6	12.03**	3.32
SWP	18	96	0.67	0.77
SWV	6	96	0.04	1.61
SPV	6	96	0.39	0.58
WPV	9	18	0.38	1.00
SWPV	18	96	1.22	1.28

Legend:
*: $p < .05$
**: $p < .01$
***: $p < .001$
S: SPEAKER
W: CONTEXTUAL VOWEL
P: PLACE
V: VOICE

speakers. The effects of PLACE-by-VOICE and VOWEL-by-PLACE were also found to be significant at .01 and .05 levels, respectively, which means that the effect due to PLACE is coupled with VOICE and VOWEL factors.

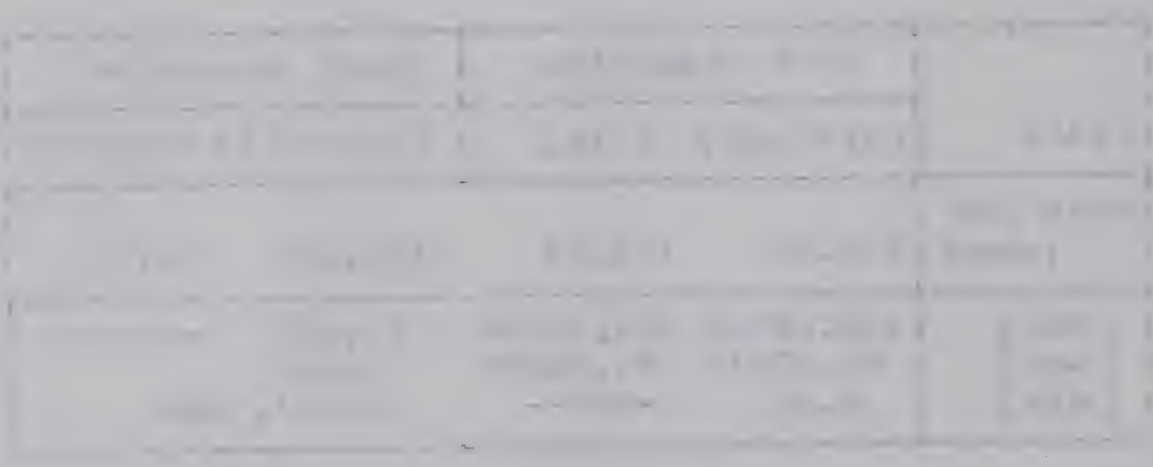
A set of Newman-Keuls tests (Winer, 1971:384-388,442) was conducted to examine the significance of the differences

of the mean duration of each individual PLACE level. The results as shown in Table 4 illustrate that there are highly significant duration differences between any fricatives belonging to different superordinate PLACE classes, while there are no or only weakly significant differences between any fricatives of the same class. It appears that [voiceless] has a longer frication than [voiced], but the

Table 4. *F*-ratios of between-PLACE differences in duration.

	back fricative		front fricative	
PLACE	[alv-pal]	[alv]	[labial]	[dental]
mean DUR (msec)	175.97	154.57	102.89	89.59
[den]	208.97***	200.37***	29.65*	------
[lab]	81.17***	75.84***	------	
[alv]	0.09	------	<df=1, 48>	

magnitude of difference is reduced for a more posterior PLACE as depicted in Figure 2. It also appeared that the effect of PLACE upon duration is slightly reduced in the context of vowels /ʌ/ and /u/ as illustrated in Figure 3. But, not only because the significance level of VOWEL-by-PLACE is relatively lower than the other effects, but also because an even higher significance of the interaction effect of SPEAKER-by-VOWEL is observed, it is presumed that the interaction of VOWEL-by-PLACE is not likely to be



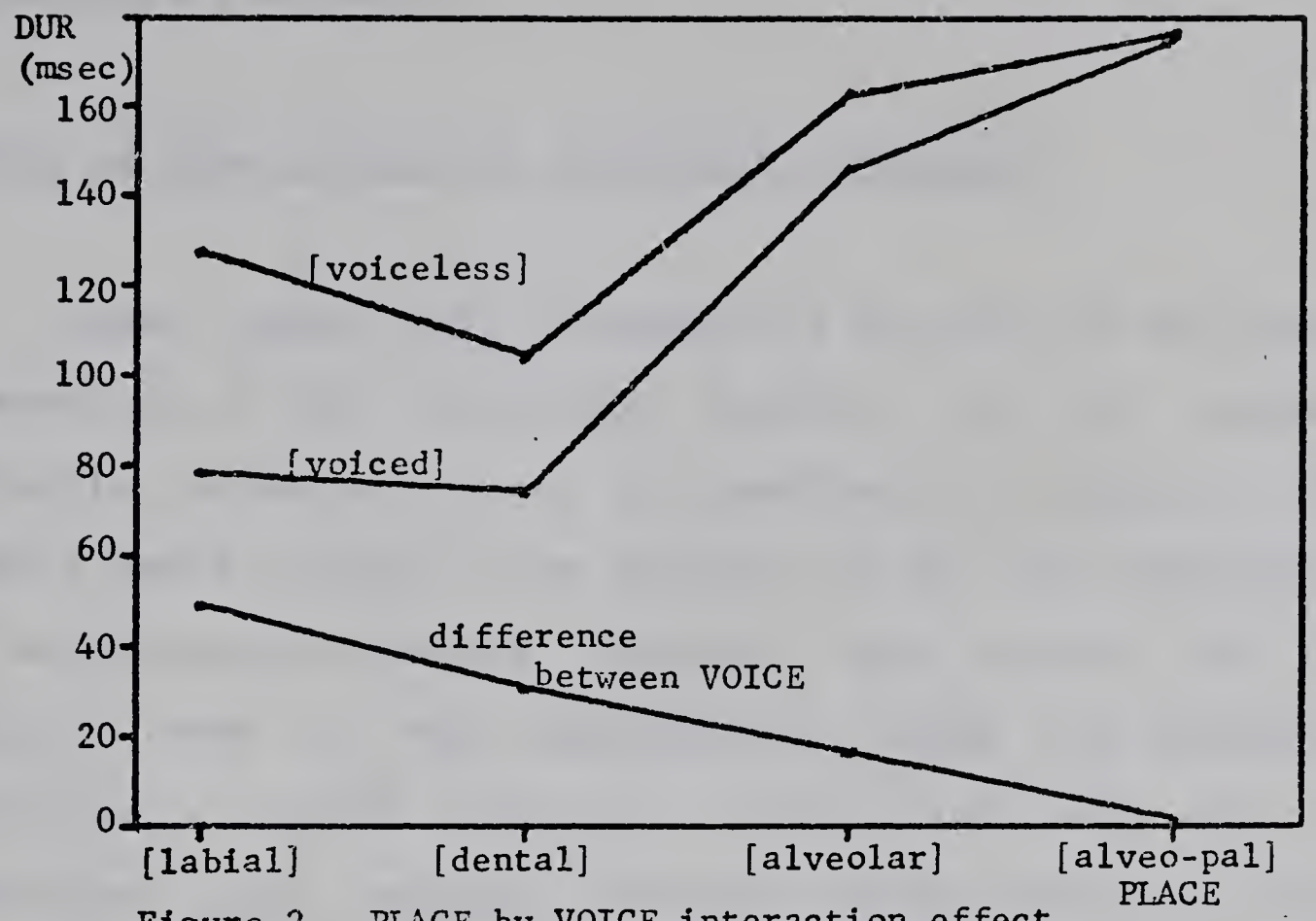


Figure 2. PLACE-by-VOICE interaction effect on duration (DUR).

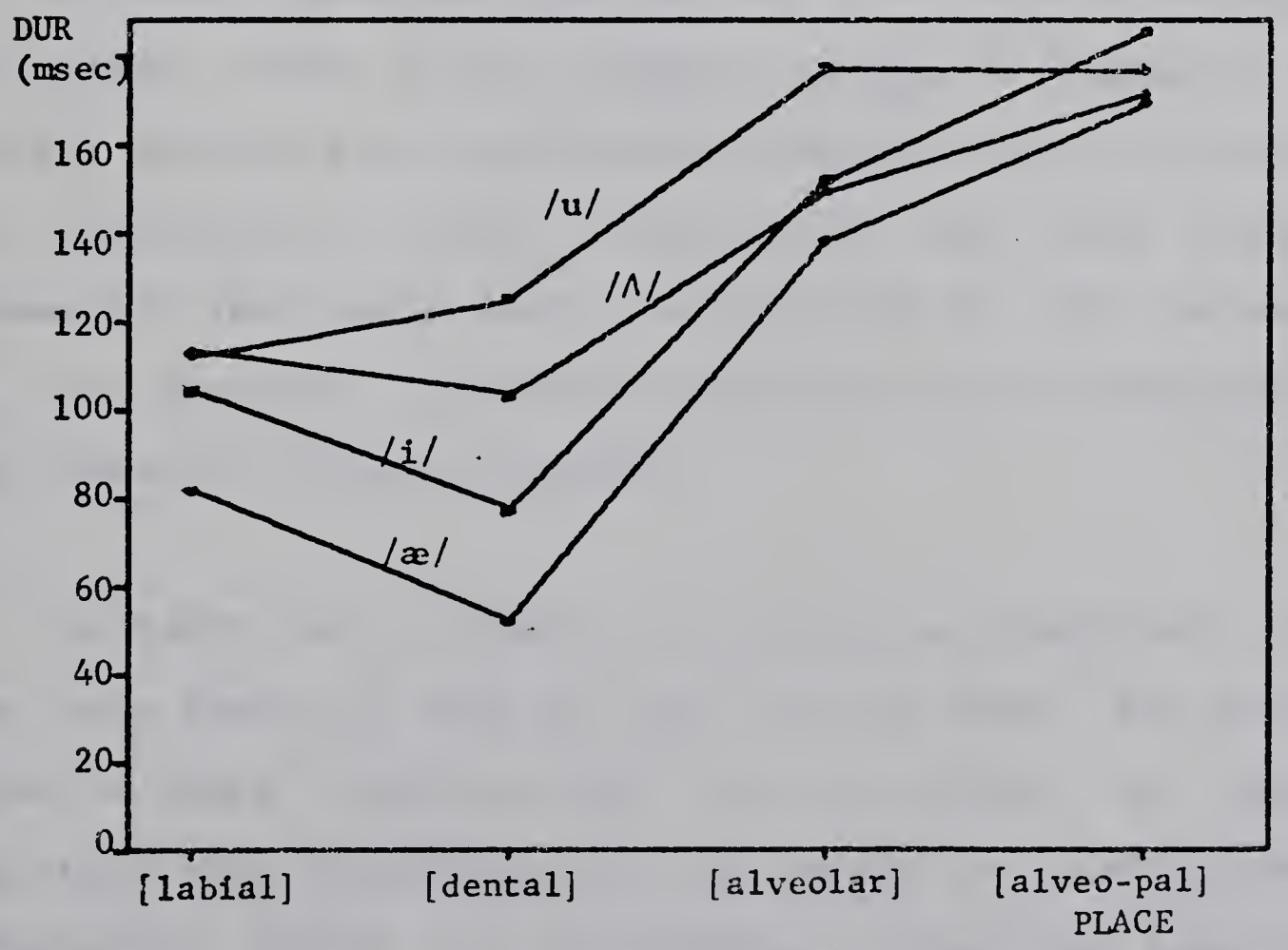


Figure 3. VOWEL-by-PLACE interaction effect on duration (DUR).

perceptually salient.

Tests of significance of intensity difference

These tests were intended to discover if the overall intensity of the frication portion (or any parameter directly dependent on it) is significantly related to PLACE and/or VOICE factors. The question of the most suitable way of measuring the overall intensity was raised for these tests. Even if the experimenter urged the speakers to maintain a stable intensity level, and kept all the pertinent gain controls constant during each step of data manipulation, the absolute value of the intensity measurement in reference to a non-contextual arbitrary level is not only apparatus-dependent but also lacks a perceptual motivation. Thus, it was decided *a priori* to normalize the overall intensity of the frication portion with reference to the contextual vowel. Henceforth, the term "overall intensity" (INT) will denote the dB value of the intensity of the frication portion in reference to the intensity of the contextual vocalic portion.

An ANOVA for the overall intensity was conducted using the same design as that for the duration test. The results shown in Table 5 indicate that the main effects of SPEAKER and PLACE were significant at .001 level, and PLACE-by-VOICE interaction effect at .01 level. It was also found that

Table 5. F-ratios of ANOVA for intensity.

Source	df		F-ratio
	num.	denom.	
S	2	96	24.56***
W	3	6	5.32*
P	3	6	99.35***
V	1	2	0.36
SW	6	96	3.01*
SP	6	96	2.30
WP	9	18	3.04*
SV	2	96	2.47
WV	3	6	0.98
PV	3	6	11.04**
SWP	18	96	0.46
SWV	6	96	1.76
SPV	6	96	0.88
WPV	9	18	1.07
SWPV	18	96	0.72

Legend:
*: p < .05
**: p < .01
***: p < .001
S: SPEAKER
W: CONTEXTUAL VOWEL
P: PLACE
V: VOICE

VOWEL, SPEAKER-by-VOWEL, and VOWEL-by-PLACE were all significant at .05 level. The high significance of both SPEAKER and PLACE suggests that, though the intensity varies with SPEAKERs, an obvious PLACE effect upon overall intensity can still be observed.

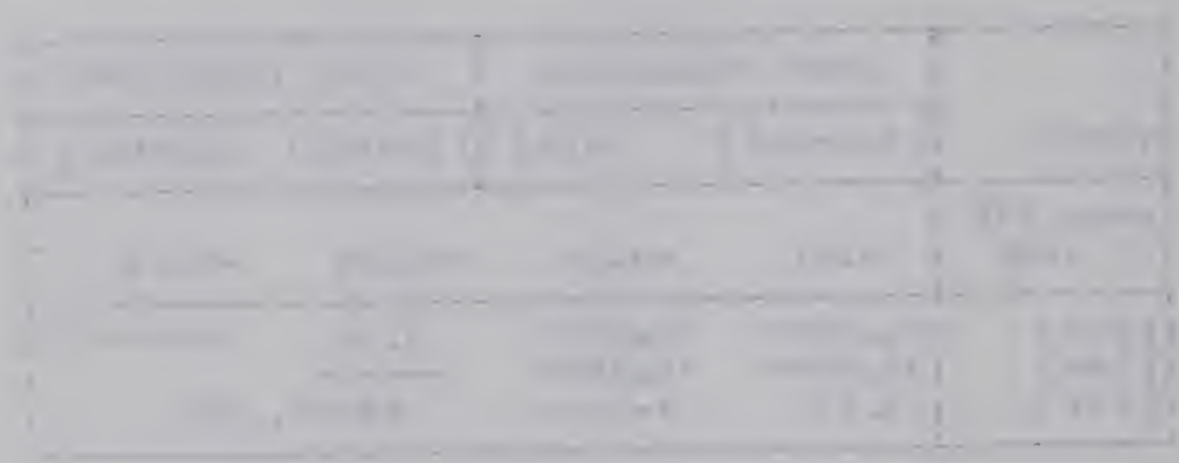
Newman-Keuls tests were conducted to examine the significance of the differences of the mean intensity of

each individual PLACE level. As shown in Table 6, the tests indicated again that the two superordinate PLACE classes (i.e., front fricatives and back fricatives) are significantly (at .001 level) different from each other in the overall intensity.

Table 6. F-ratios of between-PLACE differences in overall intensity.

	back fricatives		front fricatives	
PLACE	[alv-pal]	[alv]	[labial]	[dental]
mean INT (dB)	-3.47	-9.06	-17.99	-18.30
[den]	52.35**	29.63**	1.24	------
[lab]	37.47**	18.74**	------	
[alv]	3.21	------	<df=1, 48>	

PLACE-by-VOICE interaction as illustrated in Figure 4 indicates that front fricatives have higher intensity when voiced, while back fricatives have higher intensity when voiceless. The slight indication of the significance of VOWEL-by-PLACE interaction shows that the intensity of [alveolar] is slightly greater in the context of the rounded vowel /u/ as illustrated in Figure 5.



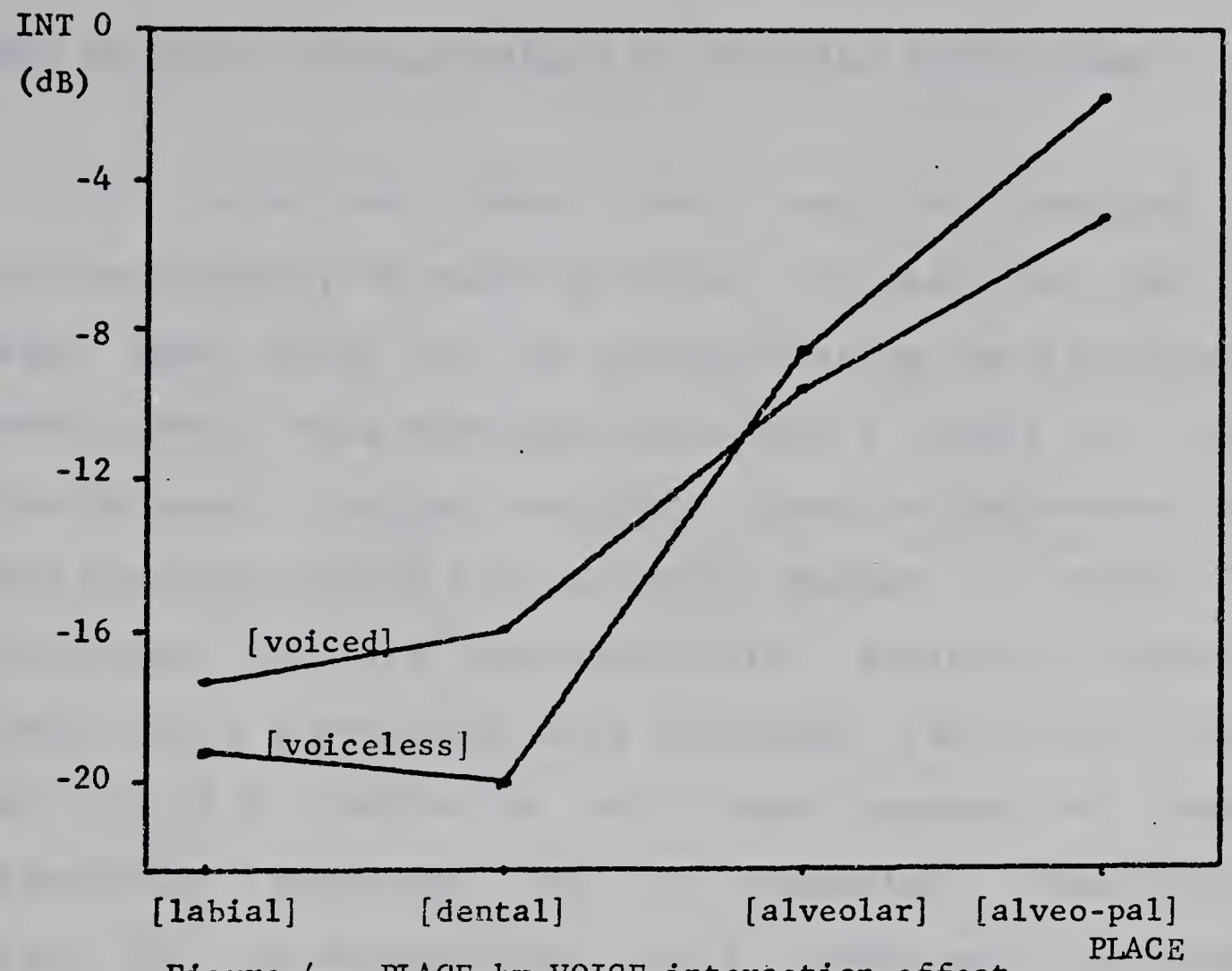


Figure 4. PLACE-by-VOICE interaction effect on intensity (INT).

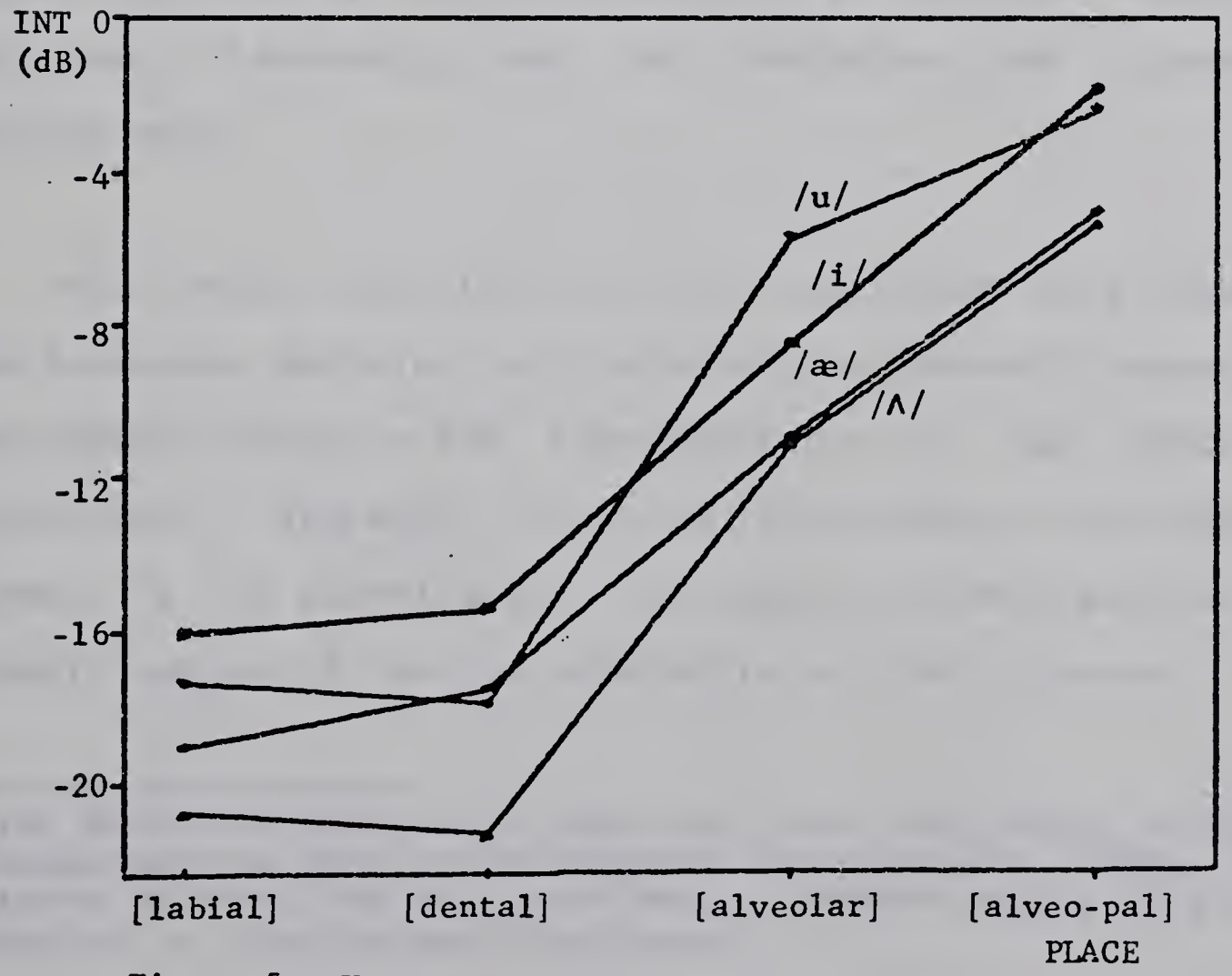


Figure 5. VOWEL-by-PLACE interaction effect on intensity (INT).

Test of PLACE discrimination by prosodic information

The aim of this test was to examine the discriminability of PLACE in terms of the two properties which were shown to be significant by the two foregoing ANOVA tests. This test was conducted by using the linear discriminant function analysis (DFA) as implemented by an SPSS package program.[1] As shown in Tables 7, PLACEs were classified 48.44 % correctly with duration information alone, 55.21 % correctly with intensity information alone, and 63.54 % correctly with both sources of prosodic information together. The corresponding identification rates of the superordinate PLACE classes were, as shown in Tables 8, 78.08 %, 86.45 %, and 92.18 %, on the basis of duration, intensity, and both duration and intensity, respectively.

The results from this test are consistent with those of the foregoing ANOVA's, and indicate that prosodic properties contribute little to the identification of the VOICE of fricatives. Evidently they do contribute, to a certain extent, to the identification of PLACE, and very much to the identification of the two superordinate PLACE classes.

---

[1] For detailed discussion about DFA, see Nie, Hull, Jenkins, Steinbrenner & Bent (1975:434-448) and Tatsuoka (1970). The options adopted for this test were a direct method with no rotation of discriminant functions.

Table 7.a Discrimination of PLACE by duration. (%)

	Predicted: [labial]	[dental]	[alveolar]	[alv-pal]
Actual:				
[lab]	22.9	47.9	25.0	4.2
[den]	6.2	62.5	20.8	10.4
[alv]	20.8	2.1	45.8	31.2
[alv-pal]	4.2	0.0	33.3	62.5

prob. of random prediction: 25.00 %
mean correct prediction: 48.44 %

Table 7.b Discrimination of PLACE by intensity. (%)

	Predicted: [labial]	[dental]	[alveolar]	[alv-pal]
Actual:				
[lab]	52.1	31.2	16.7	0.0
[den]	47.9	29.2	22.9	0.0
[alv]	0.0	14.6	66.7	18.7
[alv-pal]	0.0	0.0	27.1	72.9

prob. of random prediction: 25.00 %
mean correct prediction: 55.21 %

Table 7.c Discrimination of PLACE by duration and intensity. (%)

	Predicted: [labial]	[dental]	[alveolar]	[alv-pal]
Actual:				
[lab]	41.7	45.8	12.5	0.0
[den]	29.2	60.4	10.4	0.0
[alv]	4.2	4.2	72.9	18.7
[alv-pal]	0.0	0.0	20.8	79.2

prob. of random prediction: 25.00 %
mean correct prediction: 63.54 %

Table 8.a Discrimination of superordinate PLACE classes by duration. (%)

	Predicted: front	back
Actual: front	69.75	30.25
back	13.6	86.4
prob. of random prediction:		50.00 %
mean correct prediction:		78.08 %

Table 8.b Discrimination of superordinate PLACE classes by intensity. (%)

	Predicted: front	back
Actual: front	80.2	19.8
back	7.3	92.7
prob. of random prediction:		50.00 %
mean correct prediction:		86.45 %

Table 8.c Discrimination of superordinate PLACE classes by duration and intensity. (%)

	Predicted: front	back
Actual: front	88.55	11.45
back	4.20	95.80
prob. of random prediction:		50.00 %
mean correct prediction:		92.18 %

3. ANALYSIS OF SPECTRAL CONFIGURATIONS

This section examines the spectral characteristics of PLACE and VOICE of fricative consonants.

Normalization of spectra

The first question raised here was what was to be chosen as a reference for the intensity level of each spectrum. Three alternatives were considered:

(20)-(i) the absolute intensity level,

(ii) the level of the overall intensity of each contextual vowel, and

(iii) the maximum intensity level within each spectrum.

The absolute overall intensity must be highly correlated with the mean of the dB levels within each BIN. Hence, it was concluded that normalizing the spectral levels to any external reference level was not suitable. Therefore, the alternatives (i) and (ii) were eliminated. Accordingly, the spectra obtained by the measurement program were all re-adjusted so that each spectrum could be normalized in reference to the maximum peak within each spectrum itself.

Test of orthogonality of PLACE and VOICE

It was discussed in Chapter One that PLACE and VOICE are articulatorily independent of each other, and that, furthermore, from source-filter theory of speech production, that the acoustic effects should be essentially crthogonal. Therefore, it was decided to numerically decompose the spectra into their PLACE and VOICE components. In order to test the orthogonality between the two factors, the eight combinations of PLACE by VOICE were processed by a DFA assuming the eight were all independent non-decomposable categories. The eight categories were 86.98 % correctly identified when the prosodic information was included, and 80.73 % correctly when it was not included[1]. The centroids of each of the eight categories in the standardized space determined by the four most significant discriminant functions (out of total of seven) were examined. Figures 6 illustrates the centroids of the eight categories were arranged according to each PLACE and VOICE . The centroids of all the cognates cf different VOICE value are located very close to each other in the spaces of the first two functions as shown in Figures 6.a and 6.b, and are nearly parallel to each other in the spaces of the last two functions as illustrated in Figures 6.c and 6.d. This result indicates that, regardless how the spaces are

---

[1] Note that the probability of random prediction was 12.50 %.

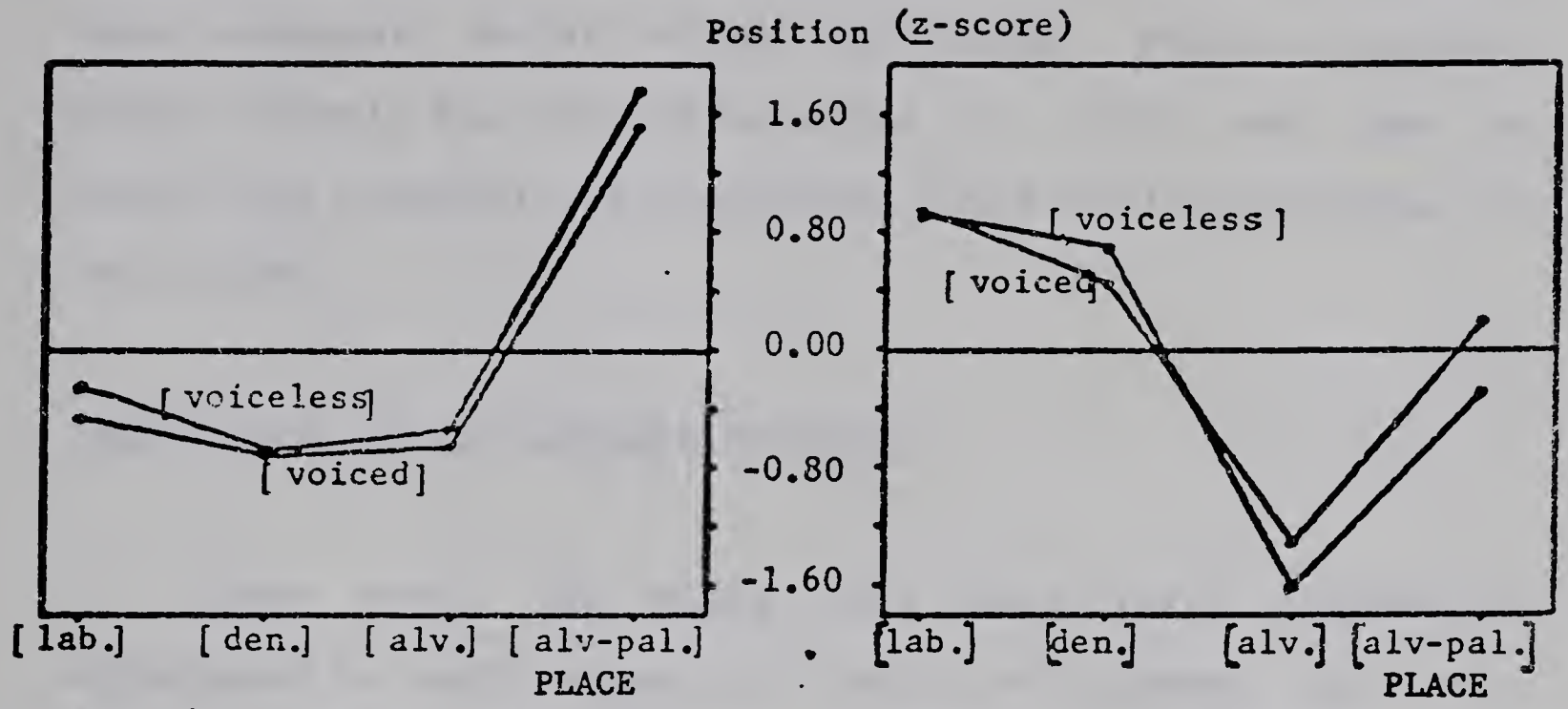


Figure 6.a Centroids of PLACE-by-VOICE categories for function 1.

Figure 6.b Centroids of PLACE-by-VOICE categories for function 2.

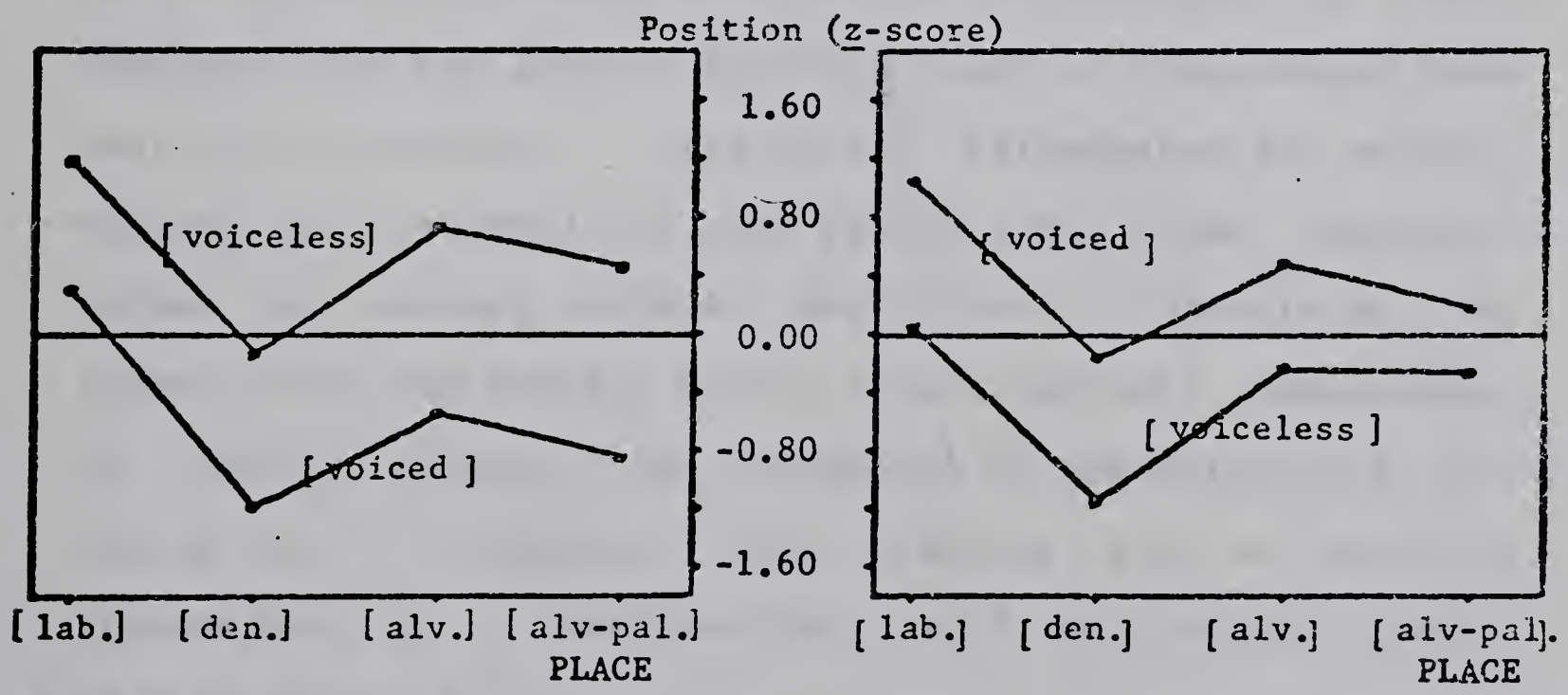


Figure 6.c Centroids of PLACE-by-VOICE categories for function 3.

Figure 6.d Centroids of PLACE-by-VOICE categories for function 4.

rotated, the centroid of each voiced PLACE category can be translated to that of voiceless cognate category by adding some constant scalar for all the PLACEs. Thus, it appears quantitatively that the effects due to VOICE and due to PLACE upon spectral configurations are largely orthogonal to each other.

Examination of the effects of VOICE

Since VOICE and PLACE have been found reasonably orthogonal to each other, it will be assumed that the difference in the spectral levels at each BIN between [voiced] and [voiceless] cognates is due to the VOICE factor. Consequently, the "voicing effect" is defined here as the signed difference obtained by subtracting at every frequency BIN the average spectral level of [voiceless] from that of [voiced].[1] Figure 7.a illustrates the voicing effects thus computed for each PLACE value (the numerical values of voicing effects are given in Appendix B). It appears that the vcicing effect is not entirely independent of PLACE. However, the curves can be approximated by the sum of the two effects. This relation may be formally represented by a mathematical model for vcicing effect, Ev(f,P) of PLACE P:

(21) $Ev(f,P) = F(f) + C(P) + e,$

---

[1] Addition of spectral levels represented in dB is equivalent to multiplication of amplitude spectra, and hence in accord with the source-filter theory (Fant, 1954).

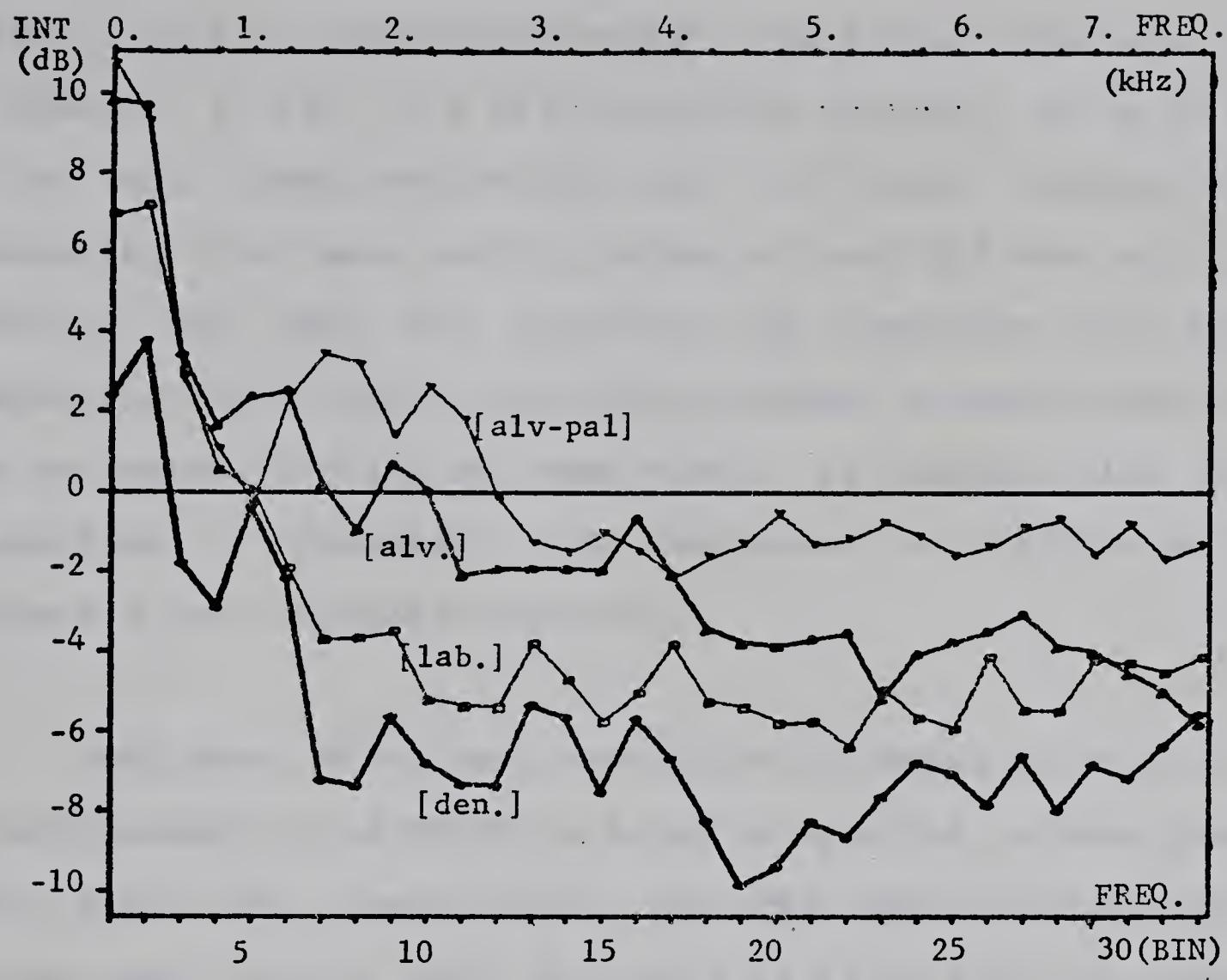


Figure 7.a Voicing effect for each PLACE.
Ev(f,P) in Equation (21).

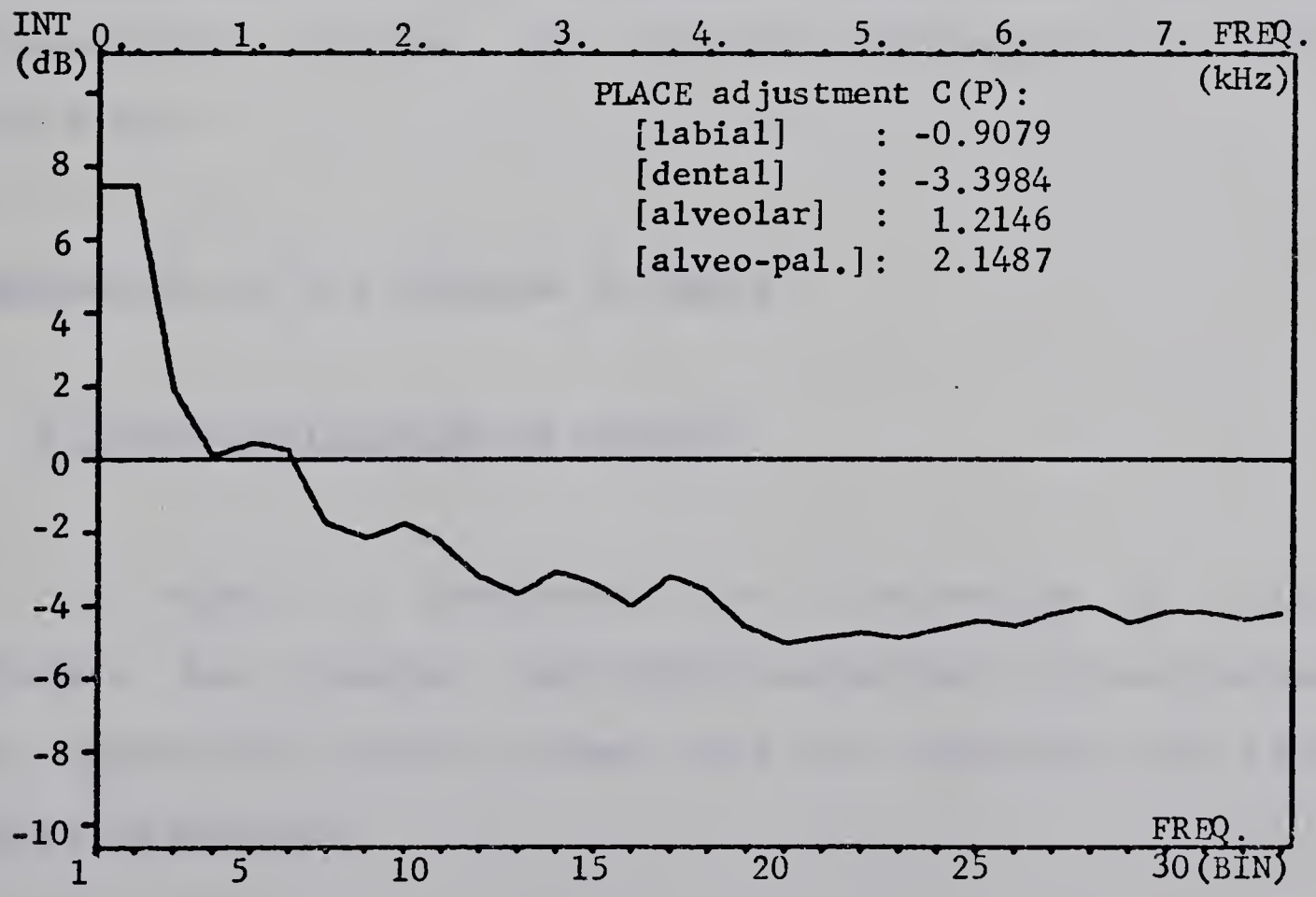
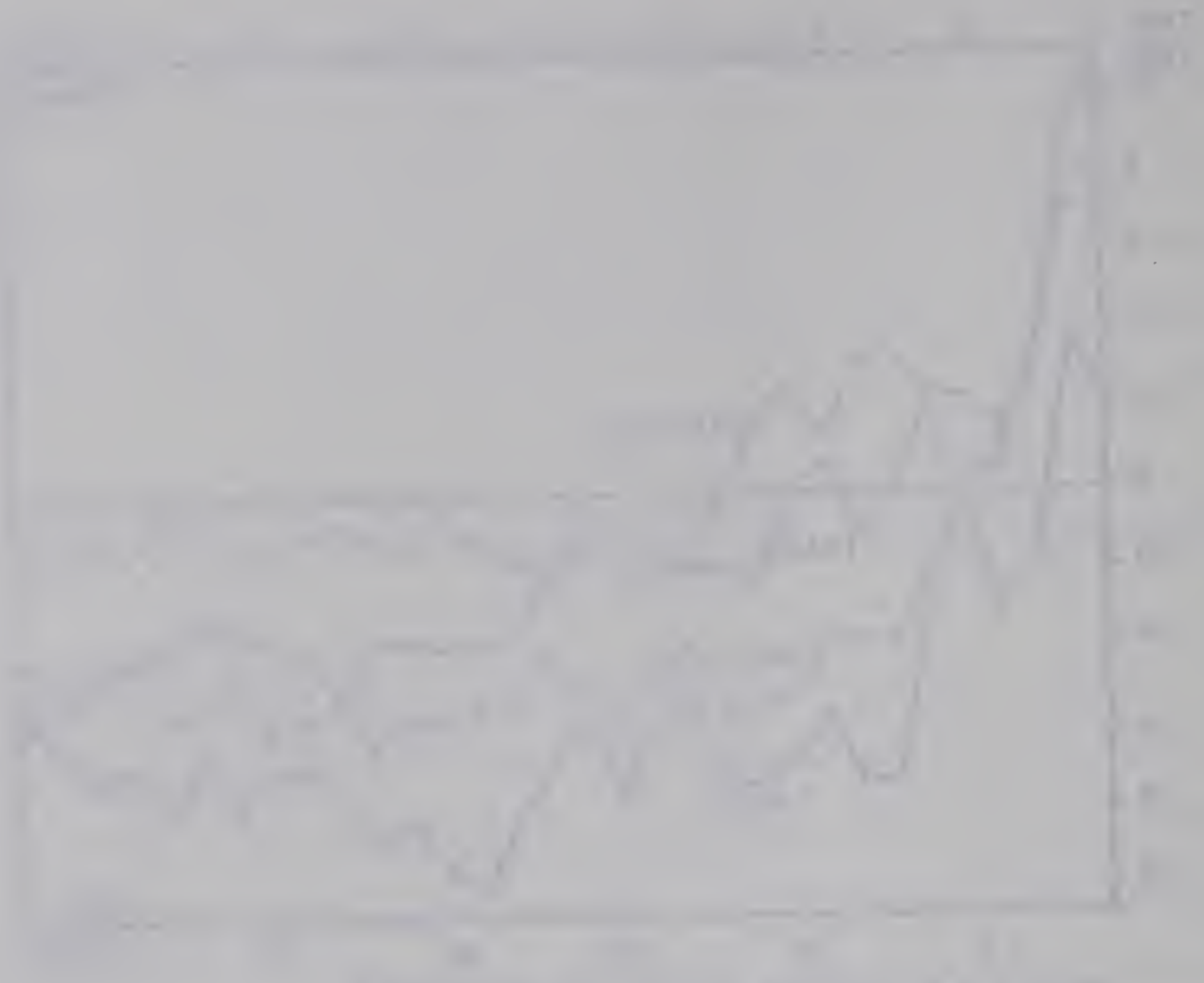
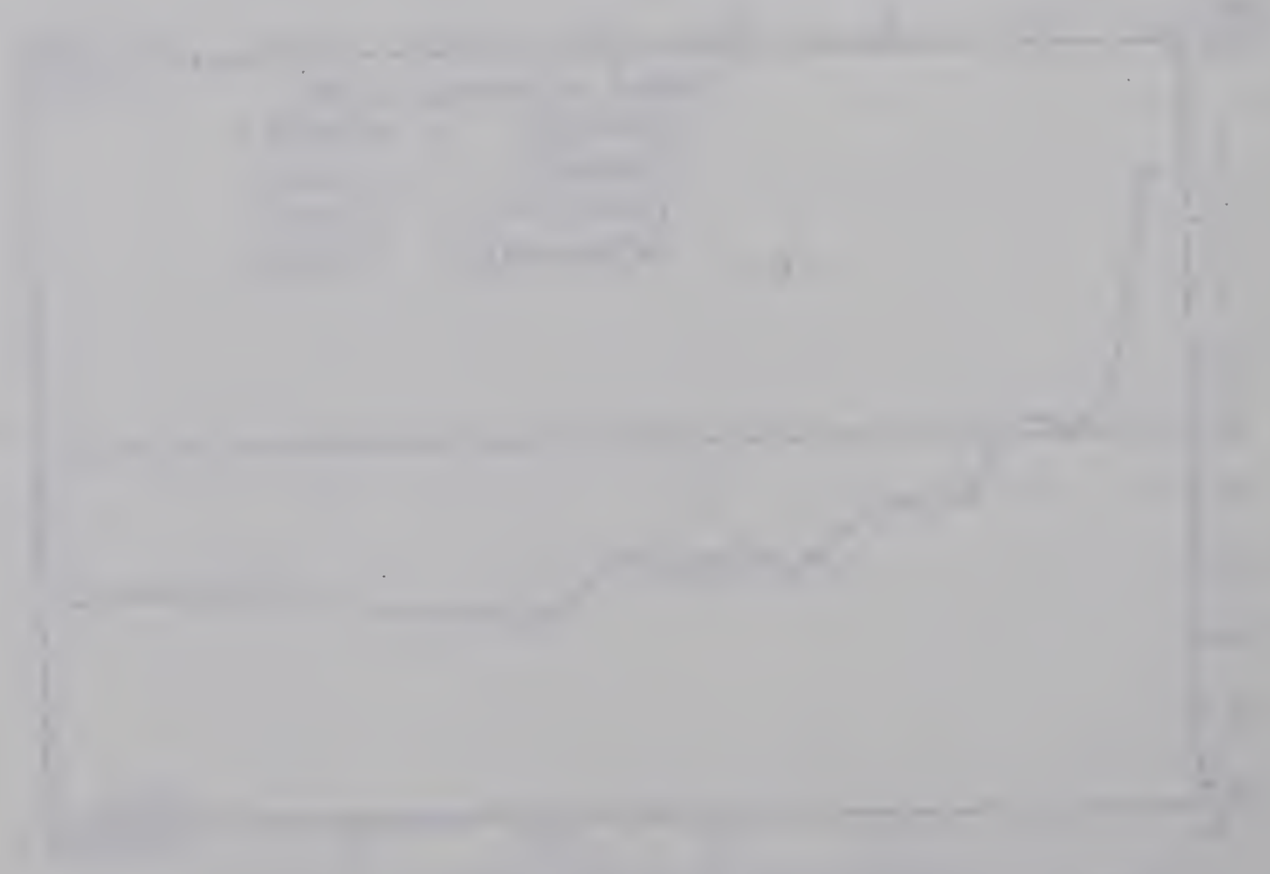


Figure 7.b Generalized voicing effect.
F(f) in Equation (21).





where F(f) is a PLACE-independent non-linear function of frequency f, C(P) is a PLACE-specific constant, and e is an error term. Estimates of F(f) and C(P) were obtained by computing the mean voicing effect at each BIN over all the PLACEs, and C(P) was estimated by computing the mean deviation from F(f) of the voicing effect of each PLACE for an estimation of C(P) for each PLACE. In general, with the exception of [dental], C(P) increases as PLACE is moved toward a more posterior position.

Additional DFA's were carried out in order to test the discriminability of VOICE in terms of spectral information. The first DFA demonstrated that the spectral information alone could predict VOICE correctly in 83.33 % of the cases. The inclusion of the prosodic information did not substantially improve the correct discrimination of VOICE (83.78 %).

Examination of the effects of PLACE

VOICE-normalization of spectra

In order to facilitate the examination of PLACE effects, the spectra were VOICE-normalized by subtracting the appropriate voicing effect from the spectrum of every voiced fricative.

Visual examination of general spectral shapes

The spectral configuration obtained after the VOICE-normalization can reasonably be thought of as purely due to the effect of PLACE. Figures 8 present the average spectrum for each PLACE. The similarity between the class members within each superordinate PLACE class is apparent. Back fricatives have pronounced spectral peaks rising 23 dB to 30 dB higher above the base. Front fricatives are characterized by a relatively flat energy distribution spread over the whole spectrum with no pronounced peaks or valleys. The spectral peaks of [alveolar] and [alveo-palatal] are found at 5.1 kHz and 3.1 kHz, respectively. The appearance of the global spectra of front fricatives are more uniformly flat with [dental] than with [labial]. The PLACE [labial] shows a broad spectral peak of about 8 dB high centered around 2.4 kHz and slight suppressions at both spectral ends.

4. PLACE DISCRIMINATION BY DFA

In order to test the discriminability of PLACE, a DFA was conducted including both the spectral information and the prosodic information. This DFA resulted in 94.79 % of correct PLACE discrimination of 192 tokens in total based on

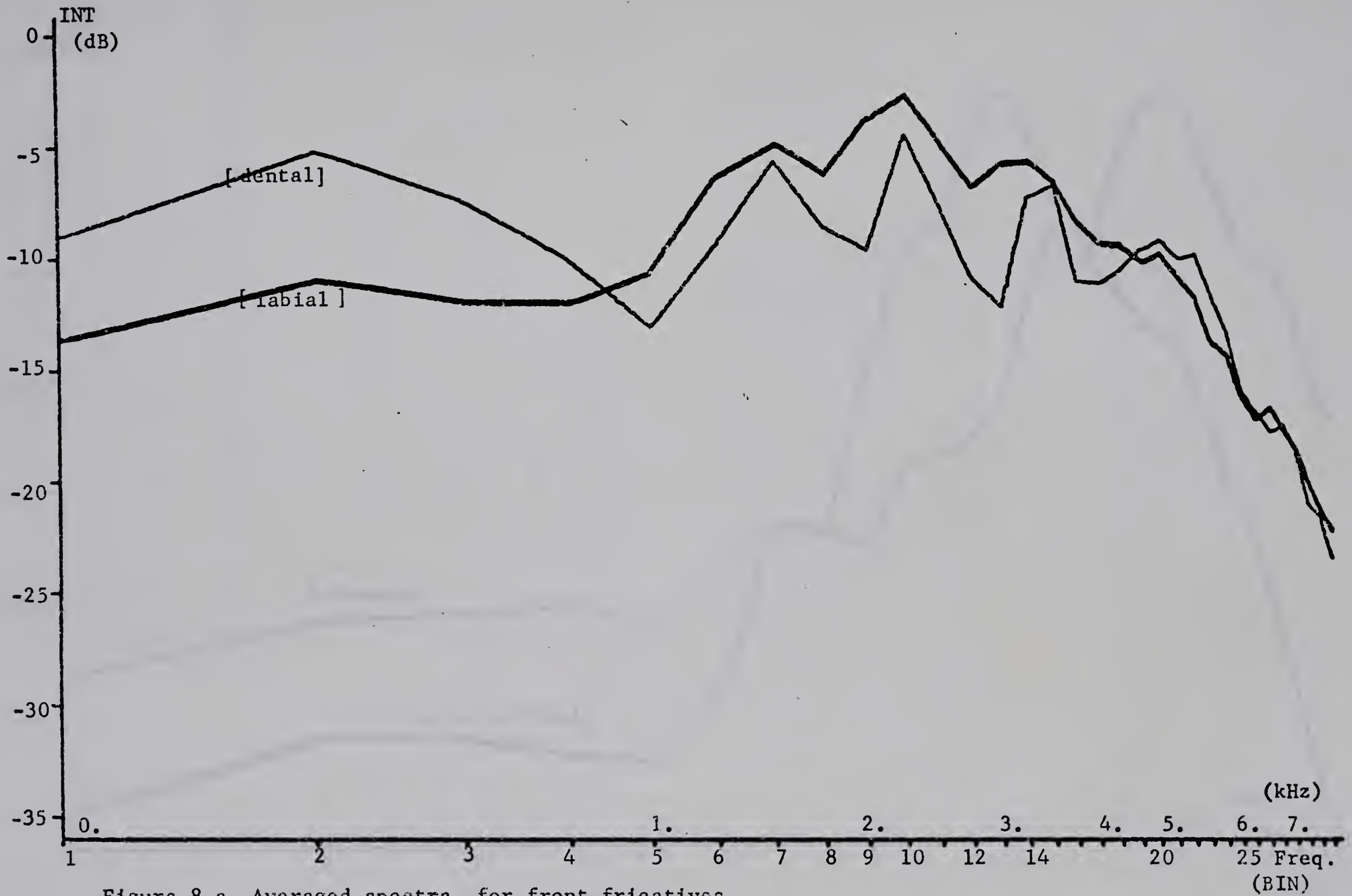


Figure 8.a Averaged spectra for front fricatives

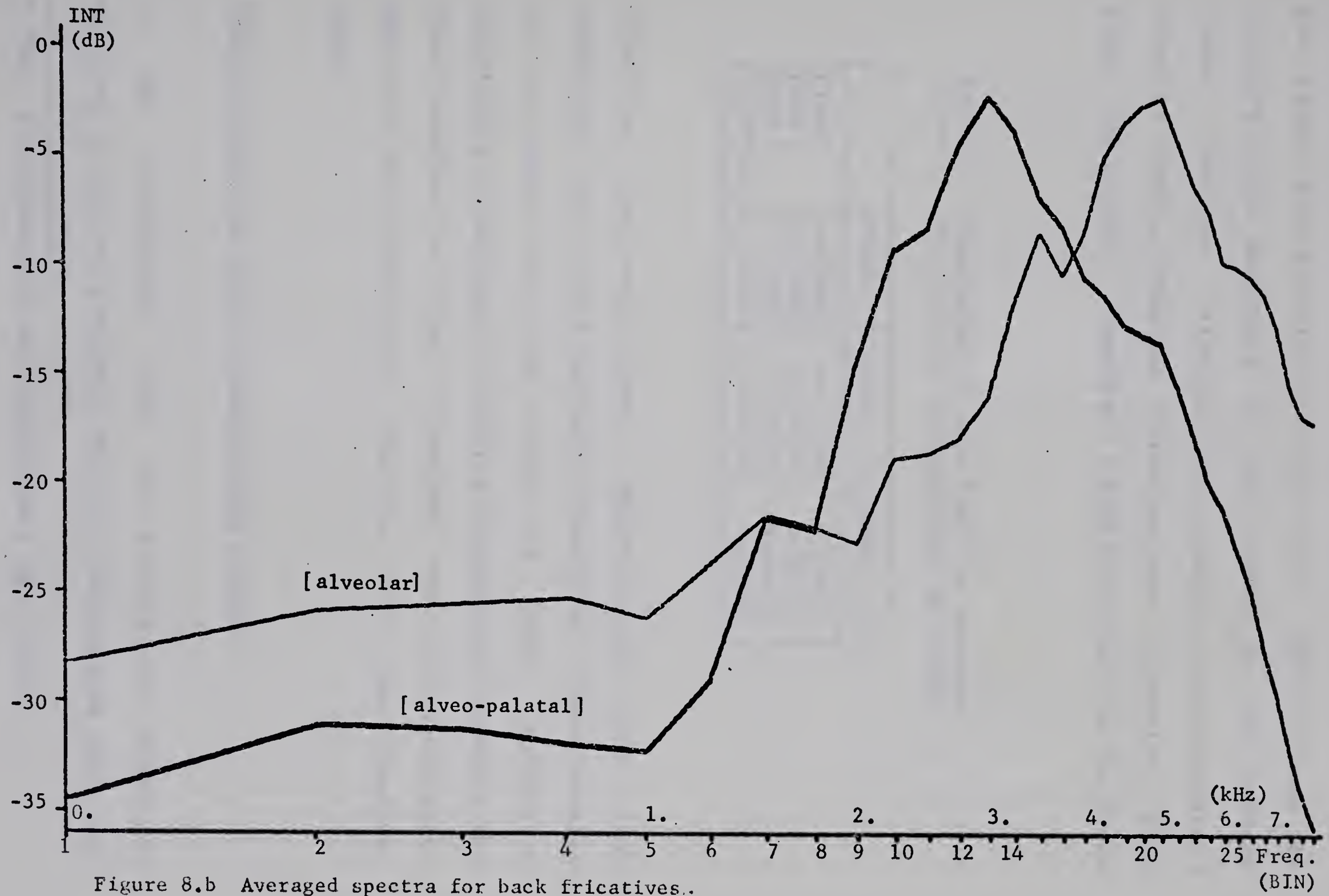


Figure 8.b Averaged spectra for back fricatives..

the three best discriminant functions[1]. Table 9 shows the location of the centroids of each PLACE in the reduced space. Function 1 distinguishes primarily [alveo-palatal] from the others. Function 2 distinguishes [alveolar] from front fricatives. Function 3 distinguishes the two front

Table 9. Centroids of each PLACE in reduced space by DFA with all information. (z score)

	Func. 1	Func. 2	Func. 3
[lab]	-0.55023	-0.89078	-0.89405
[den]	-0.76993	-0.46016	0.97761
[alv]	-0.31176	1.56875	-0.20413
[alv-pal]	1.63212	-0.21773	0.12078

fricatives from each other. Table 10 shows that back fricatives were 100 % correctly classified. These results are in accord with the reports of other perceptual experiments or analyses of English fricatives. However, the discrimination of front fricatives was also quite high, which was not anticipated from the literature (Harris, 1958).

Discrimination of superordinate PLACE classes

The coefficients for the discriminant functions obtained by the DFA involving spectral information only and

---

[1] These functions were all significant ($p < .001$).

Table 10. Discrimination of PLACE categories by all information. (%)

	Predicted: [labial]	[dental]	[alv]	[alv-pal]
Actual:				
[lab]	87.5	12.5	0.0	0.0
[den]	8.3	91.7	0.0	0.0
[alv]	0.0	0.0	100.0	0.0
[alv-pal]	0.0	0.0	0.0	100.0
prob. of random prediction:				25.00 %
mean correct prediction:				94.79 %

the averaged spectra shown in Figures 8 were examined with a view to reduce the number of spectral variables.

The averaged spectra in Figures 8 indicate that the overall spectral configuration of back fricatives and that of front fricatives are quite distinct. Furthermore all the functions with the highest canonical correlations were found to be the ones which primarily separated the two superordinate PLACE classes.[1] Thus, a number of analyses were carried out to obtain a simple formula which would capture the generality in this distinction. The main spectral source for the distinction was apparently the proportion of the energy in the high frequency and low frequency areas. The averaged spectra (Figures 8) showed

---

[1] Note that Figures 6 show that different PLACEs have significantly different values on the first two functions while different VOICEs do not.

that all four PLACEs have a slight valley at between bin 8 and bin 9 (at about 2 kHz), and the most abrupt level changes for back fricatives are in that region. So, the whole frequency scale was divided into two regions, namely, base region and upper region, with a boundary at 2 kHz at which there is an abrupt change in spectral level for the back fricatives. The mean level of each region was computed from the spectra which were obtained after VOICE-normalization. When the mean was computed, the BINs beyond bin 27 (above 6.5 kHz) were eliminated because their spectral levels were attenuated by filters during the gating process. A number of DFA's were carried out using the mean spectral levels of the two regions as shown in Table 11. The DFA revealed that 94.97 % of all superordinate PLACE classes could be correctly classified. The classification score of another DFA run with the data limited to voiceless tokens was almost perfect (98.96 %)[1]. The mean levels in the upper region are, as can be seen in Table 11, not so different between the two classes. The main source of information for the distinction between front and back fricatives lies in the difference of energy levels in the base region. Thus, another DFA including only base region was carried out, which classified the superordinate PLACE classes 92.71 % correctly. The standardized distance to the criterion function from the centroid of each category was

---

[1] Even here, the only case of the incorrect classification showed that the probability of group membership was close to chance ($p=0.618$) while all others were close to unity.

Table 11. Mean region levels of each PLACE of superordinate PLACE classes --- [voiced] signals were devoiced by VOICE-normalization. (dB)

	[labial]	[dental]	[alv]	[alv-pal]
base region	-8.8681	-9.4168	-22.9540	-27.1172
upper region	-11.6590	-13.9604	-11.5250	-12.3357

	front fricatives	back fricatives
base region	-12.8097	-25.0354
upper region	-12.8097	-11.9304

	voiceless tokens only	
	front fricatives	back fricatives
base region	-8.9213	-26.9641
upper region	-9.7064	-11.1176

computed. The mid-point was at 1.8050 standardized distance (92.82 percentile), and its raw value was -15.8929 dB. Thus, it is now claimed, with a confidence probability of 92.82 %, that a fricative consonant belongs to back fricative class if the mean level of base region is less than -15.8929 dB, and to front fricative class otherwise.

Discrimination within back fricative PLACE class

The spectral distinction between [alveolar] and [alveo-palatal] will be considered briefly in light of present data. The more difficult question of the distinction between [dental] and [labial] will be addressed in Chapter Four. A visual inspection of the averaged spectra indicates that the spectral function of [alveolar] and [alveo-palatal] cross each other twice. The first crossing was found at the previously defined boundary, i.e., at about 2 kHz. The second crossing was found between bin 16 and bin 17 (at about 4 kHz), dividing the upper region into two sub-regions: lower-upper region and higher-upper region. A DFA was conducted to test the discriminability of the two PLACEs with the spectral data obtained after voicing normalization. Except one incorrect identification, all the tokens were correctly identified (98.96 % correct identification). When the spectra of voiceless fricatives were analyzed separately, the identification rate was 100 %. The coefficient weights assigned to the variation of each spectral region was roughly 1:5:5 beginning with the lower region. This means that the discrimination was mostly based on the information residing in the upper region alone. Moreover, the standard deviation of [alveolar] tokens was 3.27 dB in lower-upper region and 1.45 dB in higher-upper region, and the standard deviation of [alveo-palatal] tokens in lower-upper region was 1.70 dB while that of the higher-

upper region was 3.09 dB. The fact that each PLACE shows a substantianlly smaller standard deviation in a region of its peak may indicate that the position and the level of their spectral peak are to some extent SPEAKER-invariant and VOWEL-invariant, as reported by LaRiviere et al. (1975). Therefore, it appears to be sufficient to check the peak frequency for distinguishing between [alveolar] and [alveo-palatal] fricatives.

CHAPTER FOUR

PERCEPTUAL EXPERIMENT AND ANALYSIS OF FRONT FRICATIVES

From previous research, as well as the present study, it appears that the physical and perceptual distinctiveness between [alveolar] and [alveo-palatal] and between the superordinate classes is clear. However, the differences between [dental] and [labial] is less well understood. The present experiment, thus concerned with the front fricatives only, has two objectives. First it addresses the long standing question of the perceptual discriminability of the PLACE of two front fricatives. The results of this experiment are then compared with those of DFA's in an attempt to examine the perceptual validity of the foregoing analyses presented in the previous chapter.

1. DESCRIPTION OF EXPERIMENT

Listeners

Eighteen students (eight males and ten females) of the

University of Alberta taking an introductory psychology course in 1978 winter session served as subjects. They were native speakers of English and had no known hearing defects. Their participation in this experiment constituted a part of their course requirement, and they were not paid for the participation.

Stimuli

From the speech materials described in the last chapter, the 48 tokens[1] of [voiceless labial] and [voiceless dental] were selected for this experiment. These stimuli were presented in two contexts: frication portion alone and frication in their original vowel contexts. We will denote each of these stimulus types (STYPEs) by "consonant-type" and "syllable-type", respectively. Thus, there were in total 96 different stimuli. Before being acoustically presented, the stumuli were multiplied by a special onset adjusting function[2] which smoothed the first 10 msec of the stimuli so that an abrupt change of acoustic pressure due to a DC bias of the signal could be eliminated.

---

[1] These were 2 CONSONANTs by 4 VOWELs by 3 SPEAKERs by 2 SPKREPs.

[2] The first 10 msec part of the function was the initial half of a cosine-squared window function, and the rest was an identity function.

Apparatus

The instruments listed below were used in this perceptual experiment. Their technical specifications were, according to the manufacturer, as follows:

(22)-(i) Minicomputer - PDP-12A (described in Chapter Three)

(ii) Audio-frequency filter - Rockland Model 1524-01 (for signal smoothing)

- Slope of frequency response: 24 dB/oct

(iii) Headphone sets - Telephonics TDH-49

- Frequency response: 30 to 6000 Hz $\pm$ 3 dB

(iv) Power amplifier - Braun AG Type CSV 250

(v) Intensity meter - Hewlett-Packard Model 3469B

Procedure

This experiment was run by an Alligator program (henceforth "test program") which presented the stimuli to subjects through headphone sets in a sound-treated phonetic laboratory. At the beginning of each run, the test program played back a pre-stored 300 Hz sine wave for calibration purposes. While the calibration tone was being played, the signal intensity for each ear was adjusted to the same level. Then, the experimenter gave the subjects brief oral instructions listed in Appendix C. Three subjects with his/her own response switch box were tested simultaneously.

The test program could be interrupted at any arbitrary moment by the experimenter who was monitoring the experiment from another switch box. The experiment began with a brief practice session of 10 syllable-type and 10 consonant-type stimuli followed by four test sessions. The test program verbally announced to the subjects at the beginning of each session which type of stimuli they were going to hear. The stimuli, passed through a desampling filter (68 to 6800 Hz), were presented to listeners at a maximum rate of one every four seconds. The experiment consisted of four sessions. In each session, there were three blocks of presentation repetitions (PRESENTREPs) of 48 stimuli in random order. For each session, the stimuli consisted of one STYPE. Nine subjects (henceforth "Group A") were presented with syllable-type stimuli in first and third sessions and consonant-type stimuli in second and fourth sessions, while the other nine subjects (henceforth "Group B") were presented the two STYPE blocks in the reverse order.

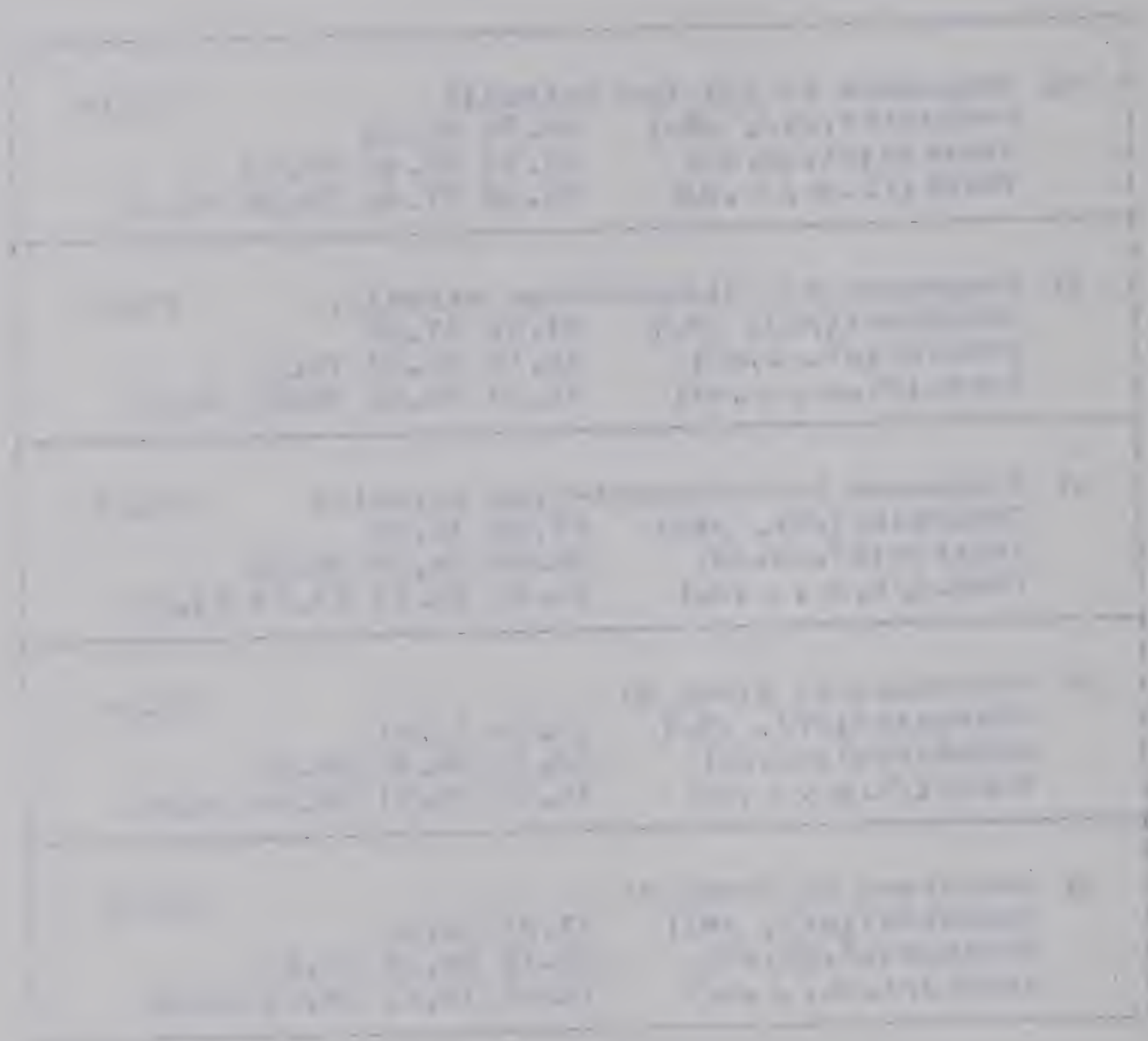
2. RESULTS AND DISCUSSION

As shown in Table 12, the total mean correct recognition rate (CRR) was about 82 % which is considerably higher than the usual significance level (75 %) for a two alternative forced-choice. Considering the fact that /f/

Table 12. Correct recognition rate. (%)

a) Responses to all the stimuli:					82.36
CONSONANT(/f/, /θ/)	89.70	75.02			
SPEAKER(s1,s2,s3)	82.52	85.82	78.73		
VOWEL(/i,æ,ʌ,u/)	76.70	77.82	79.59	95.33	
b) Responses to syllable-type stimuli:					81.83
CONSONANT(/f/, /θ/)	91.96	71.80			
SPEAKER(s1,s2,s3)	80.15	88.37	76.97		
VOWEL(/i,æ,ʌ,u/)	76.77	74.92	78.01	97.61	
c) Responses to consonant-type stimuli:					82.89
CONSONANT(/f/, /θ/)	87.54	78.24			
SPEAKER(s1,s2,s3)	84.90	83.28	80.50		
VOWEL(/i,æ,ʌ,u/)	76.62	80.71	81.17	93.06	
d) Responses by group A:					82.60
CONSONANT(/f/, /θ/)	91.59	73.61			
SPEAKER(s1,s2,s3)	82.35	86.86	78.59		
VOWEL(/i,æ,ʌ,u/)	74.77	79.01	80.56	96.06	
e) Responses by group B:					82.12
CONSONANT(/f/, /θ/)	87.81	76.43			
SPEAKER(s1,s2,s3)	82.70	84.78	78.88		
VOWEL(/i,æ,ʌ,u/)	78.63	76.62	78.63	94.60	

and /θ/ are quite frequently confused and that the hearers in this experiment were completely unable to utilize any syntactic or semantic redundancies for signal identification, the achievement of 82 % correct identification of nonsense syllables (syllable-type) or isolated noises (consonant-type), extracted from sentence context, can be regarded as highly significant. This result



indicates that front fricatives can be fairly well identified even when the frequencies above 6.8 kHz were removed[1]. This also leads to the conclusion that the energy in high frequency range is highly unlikely to be perceptually significant.

A four-way ANOVA was conducted in which STYPE, SPEAKER, CONSONANT, and VOWEL were fully crossed with two repetitions. The results shown in Table 13 indicate that SPEAKER-by-CONSONANT interaction was highly significant ($p<.001$) which implies that exemplary tokens for each front fricative consonant were produced by different speakers. The interaction effect of SPEAKER-by-CONSONANT-by-VOWEL was also significant ($p<.01$). Table 12 tells us that this interaction effect is due to the vowel /u/ which facilitates the recognition of the consonant under any condition. STYPE-by-CONSONANT interaction was also found to be significant at a lower level ($p<.05$). Concerning this interaction effect, Table 12 shows that there is an increase of CRR for [labial] and a decrease of CRR for [dental] when signals are presented in syllable-type, compared with the other stimulus type. It may be speculated that there is a response bias toward [labial] which becomes stronger when signals are more like normal speech sounds.

---

[1] Consider that the signals were filtered twice (for gating and for presentation), and more importantly the frequency characteristics of the headphones suppresses energies abruptly beyond 6 kHz.

Table 13. ANOVA for correct recognition rate

Source	df		F-ratio
	num.	denom.	
T	1	2	0.12
S	2	48	1.63
C	1	2	0.85
W	3	6	3.81
TS	2	48	0.93
TC	1	2	77.98*
SC	2	48	24.52***
TW	3	6	0.47
SW	6	48	1.94
CW	3	6	0.77
TSC	2	48	0.04
TSW	6	48	1.04
TCW	3	6	2.15
SCW	6	48	3.31**
TSCW	6	48	0.56

Legend:
*: $p < .05$
**: $p < .01$
***: $p < .001$
T: STYPE
S: SPEAKER
C: CONSONANT
W: VOWEL

Another point discovered here was that the front fricatives in front of /u/ were better identified (overall CRR = 95.33 %) under any circumstances than those in front of any other vowels. This interaction is likely to be associated with ROUNDNESS of the vowel /u/ which may contribute to affect the signal properties of the preceeding consonant. But further discussion is postponed to the last

chapter where we examine the general role of LABIALITY in modifying the signal properties.

Since it was found that there are significantly strong interactions between a number of factors which affect the correct recognition rate, an attempt was made to examine the distinctiveness of the individual tokens of front fricatives so that the quantitative analyses would be empirically compared. Firstly, the CRR's determined by SPEAKER, STYPE and CONSONANT were investigated. The results are given in Table 14. Table 14 shows that all the utterances by

Table 14. Correct recognition rates for each SPEAKER, CONSONANT and STYPE. (%)

SPEAKER	STYPE	/f/	/θ/
speaker 1	syl-type	95.72	64.49
	con-type	94.56	75.23
speaker 2	syl-type	82.18	93.56
	con-type	72.92	93.64
speaker 3	syl-type	97.69	56.25
	con-type	95.25	65.86

speaker 1 and speaker 3 are exemplary tokens for [labial], as are those by speaker 2 for [dental]. The CRR of each individual tokens was also computed so that a token-by-token comparison could be carried out later between the results of the perceptual experiment and those of the analytic

discrimination. Table 15 shows the CRR's for each token.

Table 15. Correct recognition rate for individual tokens. (%)

	/f/				/θ/			
	/i/	/æ/	/ʌ/	/u/	/i/	/æ/	/ʌ/	/u/
111	96.30	94.44	95.37	98.15	57.74	45.37	65.74	99.07
112	92.26	92.26	99.07	97.22	25.93	71.30	99.07	100.00
121	10.19	81.48	89.81	100.00	93.52	69.44	99.07	100.00
122	95.37	85.19	97.22	98.15	96.30	100.00	98.15	100.00
131	96.30	93.52	99.07	99.07	93.52	30.56	27.78	98.15
132	98.15	97.22	99.07	99.07	65.74	37.96	10.19	86.11
211	95.37	94.44	98.15	97.22	85.19	57.41	74.07	98.15
212	97.22	85.19	91.67	97.22	44.44	69.44	77.78	95.37
221	10.19	69.44	69.44	95.37	91.67	76.85	100.00	97.22
222	92.26	67.59	83.33	95.37	94.44	99.07	95.37	94.44
231	95.37	91.67	95.37	94.44	73.15	98.15	50.00	90.74
232	94.44	98.15	96.63	95.37	45.37	61.11	42.59	65.74

Note: The first three digits of each row refer to the levels of STYPE(syl-type,con-type), SPEAKER(s1,s2,s3) and SPKREP(rep1,rep2) in order.

3. PERCEPTUAL EVALUATION OF DFA'S

If the above DFA's were strictly valid as perceptual models, it would be necessary to assume that the values of the frequency BINs correspond to perceptually independent variables. A number of studies concerning signal masking (cf. Minifie et al., 1973:378-381) indicated that this

assumption is in fact too strong. However, there is as yet no consensus on these interactions among neighbouring frequency BINs for implementation in frequency analysis schemes. Hence, it will be assumed that each BIN is perceptually independent of other BINs. But the appropriateness of these assumptions can only be evaluated in light of further empirical evidence. Thus, it was decided to test the perceptual validity of DFA's (with all the 32 variables) by examining and comparing the error rate of the subjects' responses in the perceptual test and the prediction probability computed by a DFA program.

First, the ninteen tokens whose perceptual recognition was below 90 % were chosen from the perceptual experiment, and their discriminabilities by the full spectrum DFA were examined. Since VOICE was held constant as [voiceless] in the perceptual experiment, voiced front fricatives were all ignored in this comparison. Among six incorrect discriminations from the DFA, five items were found in the list of 19 perceptually worst recognized tokens. And even the one item, which was incorrectly predicted by the DFA but was not in the worst-token list, was found to have a particularly high second choice probability for the actual PLACE. Ten items in the worst-token list were predicted correctly but with a probability considerably lower than other items not in the list. However, four items in the worst-token list were well discriminated by the DFA. It was

also noticed that there was no mis-classification by the DFA of /f/ tokens uttered by speaker 1 and speaker 3 who were shown by the perceptual experiment to produce well recognized /f/ tokens. Furthermore, there was only one mis-classification of a /θ/ token uttered by speaker 2 who produced /θ/ tokens well-recognized in the perceptual experiment. With all these parallels, it seems quite reasonable to claim that the results of DFA are consistent with human performance.

4. PLACE DISCRIMINATION OF FRONT FRICATIVES

In view of the relatively high PLACE identification rates of the voiceless front fricatives by their frication portion alone, it is appropriate to examine in detail the spectral properties that distinguish them. This section attempts a discrimination of the front fricatives by means of reduced variable DFA's and then compares the results with those of perceptual recognition test.

It has been noted (Figure 8.a) that the general shape of the spectrum of [dental] is more spread while that of [labial]'s is more compact. Because of this global difference in their general shape, the spectral functions of these two PLACEs cross each other at two points. The two

cross-over points were at about 1.0 kHz and 4.3 kHz. Hence, as in the case of back fricatives mentioned in the previous chapter, three regions delimited by these crossing points were defined. With the DFA's based on the mean levels of these three regions, the two PLACEs were correctly identified at a rate of 80.21 % for VOICE-normalized tokens, and at a rate of 85.42 % of voiceless tokens only. However, the differences in mean levels of these regions were found all within one standard deviation. In the worst case, the mean level difference in the highest frequency region was only 0.9535 dB while the associated standard deviation was 5.4364 dB. Since the discriminability in each of these three broad regions is relatively weak, another approach to variable reduction was attempted. In order to search for the most predominant individual BINs in terms of discriminability, a DFA was conducted utilizing all 32 individual BINs. The identification rates[1] were 96.88 % and 100.00 % with VOICE-normalized and voiceless tokens alone, respectively. In an attempt to reduce variables, the BINs showing the highest discriminant function coefficients in the full spectrum DFA were selected:[2] bin 1 - 3 (0 - 750 Hz), bin 5 (1001 - 1250 Hz), bin 9 (2001 - 2250 Hz), bin 13 (3001 - 3250 Hz), bin 16 - 17 (3751 - 4250 Hz), and

---

[1] Notice that the identification rates by DFA are always higher than the CRR's by human listeners.
[2] This procedure may not necessarily produce an optimal set.

bin 21 - 23 (5001 - 5750 Hz).[1] REGIONs were defined where more than a single adjacent BIN among the candidates had weights of the same sign. Thus, region A was defined for 0 - 750 Hz band, region B for 3751 - 4250 Hz band, and region C for 5001 - 5750 Hz band. A DFA based on this reduced number of variables correctly classified 93.75 % of VOICE-normalized tokens and 97.92 % of voiceless tokens. The spectral levels of each PLACE were such that [labial] has more energy in all the selected BINs and region B, while [dental] has more energy in region A and region C. The discriminability of each individual variable (selected BINs or REGIONs) were examined in terms of the mean level difference in a standardized score. The difference was particularly great (distance > 1 z) in bin 9 (2001 - 2250 Hz), bin 13 (3001 - 3250 Hz), and region A. Thus, another DFA was carried out utilizing these three most prominent variables. By this DFA, 82.29 % of VOICE-normalized tokens and 85.42 % of voiceless tokens were correctly classified. Considering that the number of variables used was small, the rate seems to be highly significant. A simple comparison was made between the errors committed by the DFA with full spectral information and the last one based on the six reduced variables. All the erroneous classifications by the previous DFA were still incorrectly classified by the DFA with more reduced

---

[1] Notice that bin 5 (1001 - 1250 Hz), bin 9 (2001 - 2250 Hz), bin 13 (3001 - 3250 Hz), and region B are all the frequency areas where zeroes of [dental] exist.

variables. However, 80 % of the newly added misclassifications due to this variable reduction were associated with tokens containing contextual back vowels (0 % with /i/, 20 % with /æ/, 40 % with /ʌ/, 60 % with /u/). Considering that this reduction was primarily concerned with the energy in the 5 - 6 kHz band, it appears that the difference in the 5 - 6 kHz region is particularly related to the distinction of the PLACE of front fricatives in the context of back vowels[1]. Thus, from these analytic discrimination tests, we are led to presume that front fricatives are well classified by the spectral information in region A (0 - 750 Hz), bin 9 (2001 - 2250 Hz), and bin 13 (3001 - 3250 Hz) in general. But, if they are coarticulated with a following back vowels, then the spectral information in the 5 - 6 kHz band is also important.

---

[1] A visual inspection of indivudual spectrographs separately plotted according to the factors, VOWEL and SPEAKER confirmed that [dental] tokens uttered in the context of vowel /u/ show an important emphasis in this frequency band. Some /θ/ tokens mostly uttered by speaker 3 which induced the poorest CRR in the perceptual test clearly showed insufficient energy in this region.

CHAPTER FIVE

GENERAL DISCUSSIONS AND CONCLUSIONS

This chapter summarizes some major findings, and attempts to interpret and explain the findings in terms of an acoustic model of speech production. It examines some of their theoretical implications within the framework of recent phonetic models of speech description.

1. SUMMARY OF FINDINGS

Through a perceptual experiment, it was shown that the information contained in the frication portion alone is largely sufficient to identify PLACE of English fricatives (at least, of voiceless fricatives).[1] A set of discriminant function analyses indicated that a spectral composite can be analyzed into two largely independent components attributable to PLACE and VOICE, respectively. Based on this orthogonality, a generalized voicing effect was

---

[1] The information of formant transitions in vocalic portions could be more important for voiced fricatives.

estimated and the constants for individual PLACE-adjustments were derived. The analyses of duration, overall intensity, and spectral configuration of the frication portion indicated that PLACE categories constitute two superordinate PLACE classes, namely front and back classes.

2. GENERALIZATION OF PLACE EFFECTS

It was found that there are a number of phenomena concerning PLACE which generally hold:

(23)-(i) a more posterior PLACE is associated with a longer frication duration,

(ii) a more posterior PLACE is associated with a higher overall intensity of frication,

(iii) a more posterior PLACE has a more positive PLACE-adjustment of voicing effect, and

(iv) a more posterior PLACE has a lower spectral level in the frequency region 0 - 1000 Hz.

There are a number of possible explanations for item (ii). Differences in the nature and shapes of the constrictions may be relevant to this. It is also possible that the noise generating efficiency of more posterior fricatives is greater. Differences in pole and zero frequencies and coupling between fron and back cavities may

also be involved in this event. Item (i) also appears to be difficult to explain. However, it is possible that duration is also redundantly lengthened by the speaker for a more intense fricative. Item (iii), that a more posterior PLACE has a more positive voicing effect, seems to be related to the intrinsic intensity of the signal. It is well known that voicing generally depresses the energy in the upper region (i.e., above 2 kHz). The degree of this suppression due to voicing is likely to be more crucial for fricative signals having less overall intensity because the drop of air-pressure due to voicing can lessen the Reynold's number below the critical point more easily for a less intense signal so that no frication can take place. The fact stated in item (iv) definitely appears to be associated with the prominence of the spectral peak. It is noteworthy that there is a point consistant in many of these acoustic correlates of PLACE. The properties, duration, overall intensity, and prominence of the spectral peak, may each be used as a scale to order fricatives of different PLACEs in the same way: [dental], [labial], [alveolar] and [alveo-palatal]. Therefore, it is concluded that the signal of a more posterior PLACE has, in general, greater magnitude on each of these scales.[1]

---

[1] The results of Miller & Nicely (1955) showed that the consonants of more posterior PLACE were less confused in the perceptual test. Thus, it is suspected that the magnitude on this scale reflects the perceptual distinctivity of fricative signals.

3. CONTRIBUTION OF LABIALITY TO PLACE

A close examination of each analysis in the present study shows that the generalization we have attempted just above does not hold between [labial] and [dental]. In other words, [dental] appears to be anterior to [labial] in terms of the generalized criteria (i.e., duration, overall intensity and prominence of spectral peak), despite the fact that [dental] is articulatorily regarded as posterior to [labial]. However, this exception seems to be explained very reasonably by considering acoustic aspects of speech production. Ladefoged's (1975:270) system of feature percent value indicates that the major source for a distinction of [labial] and [dental] is rather LABIALITY than PLACE (more precisely, LOCATION).[1] In Chapter One, it was discussed that lip-protrusion can lengthen the front cavity of the vocal tract while the configuration of the back cavity remains unchanged. Thus, it is possible that the effective front cavity for [labial] could be slightly longer than that for [dental] in spite of the fact that [labial] is a more anterior to [dental] in terms of the traditional concept primarily based on articulatory LOCATION. The slight spectral rise of [labial] (compared with [dental]) centered around 2.5 kHz, might then also be

---

[1] Ladefoged assigns 90 % value of LABIALITY to /f/, and 5 % to /θ/ while assigns 95 % value of PLACE to /f/ and 90 % value to /θ/.

attributable to the difference in their LABIALITY. A similar phenomenon may take place in back fricatives as well. Because the zeroes of [alveolar] and [alveo-palatal] are found nearly at the same frequency, the length of the back cavity for these two back fricatives is not likely to be significantly different. The remarkable difference in frequency location of spectral peaks of these two back fricatives may possibly be explained by assuming that the vocal tracts for these categories differ substantially only in their front cavity.[1] It is widely observed that alveo-palatal fricatives generally involve more lip-rounding than alveolars.[2] If this observation holds generally, [labial] and [alveo-palatal] become functionally more posterior to [dental] and [alveolar] respectively although LOCATIONs may not be significantly different in their actual articulation. So, in the case of fricatives, if we regard [dental] as an extreme value of PLACE, then, PLACE is a feature more closely related to certain geometric aspects of speech production than to the anatomically defined point of articulation of traditional phonetics. Thus, it seems more likely that a feature framework based on an acoustic model of speech production would more appropriately describe the speech sounds of a language (at least, the fricatives of

---

[1] The frequencies of poles and zeroes are also subject to the shape of cross-sectional area at the constriction location.

[2] Ladefoged (1975:270) assigns to /ʃ/ 60 % value of LABIALITY and to 60 % value of ROUNDNESS and to /s/ 40 % value of LABIALITY and 40 % value of ROUNDNESS.

English). But, these speculations are not based on a sufficient experimental ground. Therefore, further studies are required to test these hypotheses in more rigorous ways.

4. DETECTION STRATEGY FOR IDENTIFICATION OF FRICATIVE PLACE

Every analytical test in the present study indicated that the two superordinate PLACE classes, namely, front and back fricatives, constitute a hierachical structure of PLACE. As to the identification of PLACE between back fricatives, a peak-detecting[1] strategy appears to work successfully. However, it was discovered in the examination of the averaged spectra of fricatives that the spectral peaks of the two front fricatives are all in the same frequency bands. Therefore, evidently a different strategy for signal identification should be required in the case of front fricatives. It appears true that the PLACE identification of front fricatives does not involve a simple strategy. In the early studies, Hughes & Halle (1956) and Jassem (1965) failed to distinguish the PLACE of front fricatives. On the other hand, Harris (1958), Heinz & Stevens (1961), Delattre et al. (1964), and LaRiviere et al. (1975) indicated the information from adjacent formant

---

[1] We will regard local maxima in the averaged spectrum as spectral peaks. Likewise, we will regard local minima in the averaged spectrum as spectral valleys.

transitions contributed strongly to the distinction PLACE of front fricatives. But, the perceptual test of the present study indicated, at least for voiceless fricatives, that:

(24)-(i) the frication portion per se of a fricative conveys sufficient information for a high rate of correct PLACE identification, and

(ii) the addition of the vocalic transition does not significantly increase the correct recognition rate.

The visual examination of the spectrum of each of the 24 tokens of front fricatives indicated that the frequency locations of zerces[1] of the two PLACEs were systematically different between the two front fricatives. A satisfactory classification was obtained by a discriminant function analysis involving a limited number of frequency bands. In these bands, there were large differences between [labial] and [dental] because of the differences in frequencies of zeroes in the spectra. Thus, it is concluded that a pole-zero detecting strategy is a plausible candidate for the modeling of the perceptual process of the PLACE identification of front fricatives.

---

[1] Zeroes were inferred from local spectral minima.

REFERENCES

Allen, W. (1953). Phonetics in Ancient India. London: Oxford Univ. Press.

Berg, J. van den (1968). Mechanism of the larynx and the laryngeal vibrations. In B. Malmberg (ed.), Manual of Phonetics, 278-308. New York: American Elserier Pub. Co.

Broad, D. & R. Fertig (1970). Formant Frequency trajectories in selected CVC nuclei. Journ. Acoust. Soc. America 47: 1571-1582.

Catford, J. (1968). The articulatory possibilities of man. In B. Malmberg (ed.), Manual of Phonetics, 309-333. New York: American Elsevier Pub. Co.

Chiba, T. & M. Kajiyama (1941). The Vowel, its Nature and Structure. Tokyo: Tokyo-Kaiseikan Publishing Co.

Chomsky, N. (1964). Current issues in linguistic theroy. In J. Fodor & J. Katz (eds.), The Structure of Language: Readings in the Philosophy of Language, 50-118. Englewood Cliffs: Prentice-Hall.

Chomsky, N. & M. Halle (1968). The Sound Pattern of English. New York: Harper & Row.

Cole, R. & W. Cooper (1975). Perception of voicing in English affricates and fricatives. Journ. Acoust. Soc. America 58: 1280-1287.

Delattre, P., A. Liberman & F. Cooper (1964). Formant transitions and loci as acoustic correlates of place of articulation in American fricatives. Studia Linguistica 16: 104-121.

Denes, P. & E. Pinson (1963). The Speech Chain. Bell Telephone Laboratories.

Fant, G. (1956). On the predictability of formant levels and spectrum envelopes from formant frequencies. In M. Halle, H. Lunt & H. MacLean (eds.), For Roman Jakobson: 109-120. The Hague: Mouton.

Fant, G. (1962). Descriptive analysis of the acoustic aspects of speech. Logos 5: 3-17.

Flanagan, J. (1972). Speech Analysis Synthesis and Perception. New York: Springer-Verlag.

Fujisaki, H. & O. Kunisaki (1978). Analysis, recognition, and perception of voiceless fricative consonants in Japanese. IEEE Trans. on ASSP 26: 21-27.

Halle, M. (1964). On the basis of phonology. In J. Fodor & J. Katz (eds.), The Structure of Language: Readings in the Philosophy of Language, 324-333. Englewood Cliffs: Prentice-Hall.

Halle, M., G. Hughes & J.-P. Radley (1957). Acoustic properties of stop consonants. Journ. Acoust. Soc. America 29: 107-116

Harris, K. (1958). Cues for discrimination of American English fricatives in spoken syllables. Language and Speech 1: 1-7.

Heinz, J. & K. Stevens (1961). On the properties of voiceless fricative consonants. Journ. Acoust. Soc. America 33: 589-596.

Hughes, G. & M. Halle (1956). Spectral properties of fricative consonants. Journ. Acoust. Soc. America 28: 303-310.

Huh, W. (1968). Korean Phonology. Seoul: Chung-Um Publishing Co.

Jakobson, R., G. Fant. & M. Halle (1969). Preliminaries to Speech Analysis (9th ed.). Cambridge: MIT Press.

Jassem, W. (1965). The formants of fricative consonants. Language and Speech 8: 1-16.

Jones, D. (1960). An Outline of English Phonetics. Cambridge, England: Heffer & Sons Ltd.

Ladefoged, P. (1962). Elements of Acoustic Phonetics. Chicago: The University of Chicago Press.

Ladefoged, P. (1967). Three Areas of Experimental Phonetics. London: Oxford University Press.

Ladefoged, P. (1971). Preliminaries to Linguistic Phonetics. Chicago: Univ. of Chicago Press.

Ladefoged, P. (1975). A Course in Phonetics. New York: Harcourt Brace Jovanovich, Inc.

LaRiviere, C. (1974). Speaker identification from turbulent portion of fricatives. Phonetica 29: 246-252.

LaRiviere, C., H. Winitz & E. Herriman (1975). The distribution of perceptual cues in English prevocalic fricatives. Journ. Speech and Hearing Research 18: 613-622.

Lehiste, I. & G. Peterson (1959). Vowel amplitude and phonemic stress in American English. Journ. Acoust. Soc. America 31: 428-435.

Lieberman, P. (1968). Intonation, perception, and Language. Cambridge: MIT Press.

Lieberman, P. (1977). Speech Physiology and Acoustic Phonetics. New York: MacMillan Pub. Co.

Lindblom, B. & A. de Serpa-Leitão (1969). Consonant confusions in English and Swedish. In G. Fant (ed.), Speech Sounds and Features: 100-109.

Massaro, D. & M. Cohen (1976). The contribution of fundamental frequency and voice onset time to the /zi/-/si/ distinction. Journ. Acoust. Soc. America 60: 704-717.

Miller, G. & P. Nicely (1955). An analysis of perceptual confusions among some English consonants. Journ. Acoust. Soc. America 27: 338-352.

Minifie, F. & T. Hixon & F. William (1973). Normal Aspects of Speech, Hearing, and Language. Englewood Cliffs: Prentice-Hall, Inc.

Müller, J. (1840). Handbuch der Physiologie des Menschen. Bd. II, Coblenz.

Nearey, T. (1977). Phonetic feature systems for vowels. Ph. D. dissertation at the Univ. of Connecticut. Mimiograph distributed by the Indiana Univ. Linguistic Club, Bloomington, Ind. 1978.

Nie, N., C. Hull, J. Jenkins, K. Steinbrenner & D. Bent (1975). SPSS: Statistical Package for the Social Science (2nd ed.). New York: McGraw Hill Book Co.

Peterson, G. & I. Lehiste (1960). Duration of syllable nuclei in English. Journ. Acoust. Soc. America 32: 693-703.

Potter, R., G. Kopp & H. Kopp (1966). Visible Speech. New York: Dover Publications, Inc.

Rabiner, L. (1967a). Speech synthesis by rules: an acoustic domain approach. Bell System Tech. Journ. 47: 17-37.

Rabiner, L. (1967b). Digital formant synthesizer for speech-synthesis studies. Journ. Acoust. Soc. America 43: 822-828.

Rabiner, L. & B. Gold (1975). Theory and Application of Digital Signal Processing. Englewood Cliffs: Prentice-Hall.

Rozsypal, A. (1976). Digital gating of speech signals. Language and Speech 19: 57-74.

Saussure, F. de (1969). Cours de Linguistique Générale. published by C. Bally & A. Sechehaye. Paris: Payot.

Stevens, K. (1971). Airflow and turbulence noise for fricative and stop consonants: static considerations. Journ. Acoust. Soc. America 50: 1180-1192.

Stevens, K. & A. House (1955). Development of a quantitative description of vowel articulation. Journ. Acoust. Soc. America 27: 484-493.

Strevens, P. (1960). Spectra of fricative noise in human speech. Language and Speech 3: 32-49.

Stevenson, D. & R. Stephens (1978a). A programming system for phychoacoustic experimentation. Paper presented at 11th DECUS Canada Symposium in Ottawa.

Stevenson, D. & R. Stephens (1978b). The Alligator reference manual. Unpublished manuscript.

Tatsuoka, M. (1970). Selected Topics in Advanced Statistics: An elementary approach No. 6: Discriminant Analysis. Champaign: Institute for Personality and Ability Testing.

Winer, B. (1971). Statistical Principles in Experimental Design (2nd ed.). New York: McGraw-Hill.

APPENDIX A

Frequency range and mid-point (in Hz) of each BIN

BIN #	from	to	mid-pt
1		250	125
2	251	500	375
3	501	750	625
4	751	1000	875
5	1001	1250	1125
6	1251	1500	1375
7	1501	1750	1625
8	1751	2000	1875
9	2001	2250	2125
10	2251	2500	2375
11	2501	2750	2625
12	2751	3000	2875
13	3001	3250	3125
14	3251	3500	3375
15	3501	3750	3625
16	3751	4000	3875
17	4001	4250	4125
18	4251	4500	4375
19	4501	4750	4625
20	4751	5000	4875
21	5001	5250	5125
22	5251	5500	5375
23	5501	5750	5625
24	5751	6000	5875
25	6001	6250	6125
26	6251	6500	6375
27	6501	6750	6625
28	6751	7000	6875
29	7001	7250	7125
30	7251	7500	7375
31	7501	7750	7625
32	7751	8000	7875

APPENDIX B

Voicing effects

BIN #	[labial]	[dental]	[alveolar]	[alv-pal]	mean
1	6.9448	2.4237	9.8235	10.7601	7.4880
2	7.1997	3.7661	9.7269	9.3339	7.5066
3	3.0406	-1.7892	3.4820	2.8151	1.8871
4	0.5502	-2.9025	1.6746	1.0812	0.1009
5	0.0601	-0.3732	2.4119	-0.2683	0.4576
6	-1.9052	-2.1736	2.5571	2.4008	0.2198
7	-3.7043	-7.2033	0.2132	3.4569	-1.8094
8	-3.6515	-7.3591	-0.9407	3.1826	-2.1922
9	-3.4760	-5.6376	0.6963	1.4265	-1.7477
10	-5.2077	-6.7852	0.0687	2.6316	-2.3231
11	-5.3560	-7.3337	-2.0850	1.8200	-3.2387
12	-5.3905	-7.3651	-1.9073	-0.1918	-3.7137
13	-3.7984	-5.3232	-1.8823	-1.2547	-3.0646
14	-4.7113	-5.6471	-1.9080	-1.4928	-3.4398
15	-5.7550	-7.4908	-1.9502	-1.0222	-4.0545
16	-5.0212	-5.6832	-0.6073	-1.4801	-3.1979
17	-3.8049	-6.6841	-2.0522	-2.1241	-3.6663
18	-5.2629	-8.2507	-3.4020	-1.6073	-4.6307
19	-5.3956	-9.8623	-3.7662	-1.2856	-5.0774
20	-5.7923	-9.4002	-3.8287	-0.5271	-4.8871
21	-5.7342	-8.2653	-3.6479	-1.3098	-4.7393
22	-6.3864	-8.6668	-3.4970	-1.1765	-4.9317
23	-4.9465	-7.6340	-5.1665	-0.7551	-4.6255
24	-5.6615	-6.7564	-4.0309	-1.0977	-4.3866
25	-5.9324	-7.0612	-3.7166	-1.5919	-4.5755
26	-4.1313	-7.8292	-3.4578	-1.3388	-4.1893
27	-5.4557	-6.6057	-3.0065	-0.8680	-3.9840
28	-5.4719	-7.9754	-3.8602	-0.6539	-4.4903
29	-4.2154	-6.8365	-3.9934	-1.5758	-4.1553
30	-4.2858	-7.1613	-4.5298	-0.7967	-4.1934
31	-4.5224	-6.3668	-4.9992	-1.6790	-4.3918
32	-4.1177	-5.5618	-5.7982	-1.3219	-4.1999

APPENDIX C

Instructions for the perceptual experiment

In the present experiment, I am interested in how people identify the two specific consonants of English, namely /f/ as in "fat" and /θ/ as in "thick". What I would like you to do is to indicate whether the each given stimulus sounds closer to /f/ or /θ/ by pressing one of the two buttons of the switch box located in front of you, which is connected to the computer. The consonants will be presented to you either alone or followed by a contextual vowel. In the latter case, you have to pay attention to the first sound in making your decision. There will be four sessions in the experiment. All the stimuli in each session are of one of the two types just mentioned. At the beginning of each session, you will be told what type of stimuli you will hear in the session.

Stimuli will come from a computer located in a separate room. The computer will not present the next stimulus until it receives the responses from all of you for each given item. It reads in your responses very quickly. But it may take a little longer. So, please keep pressing the button until you hear a very weak and short pulse which signals that the computer has read in your response. You can use as much time as you would like to in making your decision. However, please do not keep pressing the button unnecessarily long since the computer will not proceed to the next stimulus unless all the buttons are released.

If you have any questions, please ask them now. If you do not have any, please wear the headphone set given to you properly so that they cover your ears completely.